FROM SPORULATION TO INTRACELLULAR OFFSPRING PRODUCTION: EVOLUTION OF THE DEVELOPMENTAL PROGRAM OF *EPULOPISCIUM*

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

David Alan Miller

January 2012

© 2012 David Alan Miller

FROM SPORULATION TO INTRACELLULAR OFFSPRING PRODUCTION: EVOLUTION OF THE DEVELOPMENTAL PROGRAM OF *EPULOPISCIUM*

David Alan Miller, Ph. D.

Cornell University 2012

Epulopiscium sp. type B is an unusually large intestinal symbiont of the surgeonfish *Naso tonganus*. Unlike most other bacteria, *Epulopiscium* sp. type B has never been observed to undergo binary fission. Instead, to reproduce, it forms multiple intracellular offspring. We believe this process is related to endospore formation, an ancient and complex developmental process performed by certain members of the Firmicutes. Endospore formation has been studied for over 50 years and is best characterized in *Bacillus subtilis.* To study the evolution of endospore formation in the Firmicutes and the relatedness of this process to intracellular offspring formation in *Epulopiscium*, we have searched for sporulation genes from the *B. subtilis* model in all of the completed genomes of members of the Firmicutes, in addition to *Epulopiscium* sp. type B and its closest relative, the spore-forming *Cellulosilyticum lentocellum*. By determining the presence or absence of spore genes, we see the evolution of endospore formation in closely related bacteria within the Firmicutes and begin to predict if 19 previously characterized non-spore-formers have the genetic capacity to form a spore. We can also map out sporulation-specific mechanisms likely being used by *Epulopiscium* for offspring formation. Lastly we focus on the relevance of one sporulation homolog found in *Epulopiscium*, *spoIIE*, by using reverse transcription quantitative PCR to determine that its expression profile during offspring formation is highly similar to that seen in *B. subtilis*.

BIOGRAPHICAL SKETCH

David Alan Miller Jr. was born on March 23, 1983 to David A. Miller Sr. and Christina M. Miller in Amsterdam, New York and grew up in Charleston Four Corners, New York. He graduated from Fonda-Fultonville High School in June 2001 and began attending Hartwick College in Oneonta, New York the following fall.

At Hartwick, David gained interest in research. In the summer of 2003, he worked with Dr. Mary Allen at Hartwick examining the microbial communities in the fluid of pitcher plants. In the summer of 2004, he attended a Research Experience for Undergraduates (REU) program at the University of Georgia, and he worked with Dr. Anne Summers examining large plasmids in Gram-positive bacteria isolated from chicken litter. During his senior year, David began his thesis project identifying antibiotic resistant bacteria isolated from the Susquehanna River near a Oneonta wastewater treatment facility outflow pipe and examined their potential for horizontal gene transfer of antibiotic resistance genes.

In the fall of 2005, David began attending graduate school at Cornell University. He was immediately attracted to the research of Dr. Esther R. Angert, which focused on the highly unusual bacterium, *Epulopiscium*. David joined Dr. Angert's lab in the spring of 2006 and researched the evolutionary relationship of endospore formation in the Firmicutes and intracellular offspring production in *Epulopiscium* sp. type B.

Following completion of his Ph.D., David is interested in either post-doctoral research or a teaching position at a small liberal arts college.

To my parents, for encouraging me to pursue my dreams

ACKNOWLEDGMENTS

A special thank you to my advisor, Dr. Esther R. Angert. This dissertation has only been possible because of her continued support and guidance. I would also like to thank my committee members, Dr. John D. Helmann and Dr. Joseph E. Peters. They have provided valuable insight and observations that have shaped much of the research in this dissertation.

It has been a pleasure to work with everyone in the Department of Microbiology. I am extremely grateful for the friendship and advice I have received over the years from the members of the Angert lab. I would like to thank the Helmann lab, Ahmed Gaballa in particular, for helping me with many *Bacillus* techniques. Dr. Ian Hewson was instrumental in getting the genomic analyses running.

I would like to thank our collaborators. Chapter 1 would not have been possible without the help of Dr. Garret Suen. The staff of Lizard Island, Dr. Howard Choat and Dr. Kendall Clements for surgeonfish and *Epulopiscium* sample collection, making research of this fascinating organism possible.

Lastly, I thank all my friends. They have made grad school a very positive experience and have given me a number of wonderful memories.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

A COMPARATIVE GENOMIC ANALYSIS OF THE EVOLUTION OF ENDOSPORE FORMATION

ABSTRACT

Endospore formation in the Firmicutes is useful for dispersal and surviving environmental extremes, and is best studied in the Bacilli, specifically *Bacillus subtilis*. The *B. subtilis* model is an evolved version of this likely ancient sporulation program. To study the evolution of endospore formation, we searched for homologs of known sporulation genes, as defined in *B. subtilis*, in all currently sequenced genomes of the Firmicutes and created a sporulation gene profile for each organism. A tree generated by comparing these profiles clearly separates those organisms that form an endospore from those that cannot. Within each cluster, closely related bacteria, based on 16S rRNA gene phylogeny, group together, showing the lineage-specific evolution of endospore formation. We were able to identify 19 bacteria considered to be non-sporulating that grouped with the endospore-formers. By further analyzing the endospore-formers and these potential spore-formers using a core subset of sporulation genes, we identified 61 genes conserved in at least 90% of the endospore-formers and predict that seven of the potential spore-formers have the genetic capacity for sporulation. Based on its sporulation gene profile, *Clostridium sticklandii* DSM 519 grouped with the non-sporulating bacteria. Moreover, the absence of key sporulation genes suggests that *C. sticklandii* no longer has the ability to form a spore.

INTRODUCTION

Certain members of the Firmicutes can produce an endospore to escape harsh environmental conditions. Endospore-formers are found in both the Class Bacilli and the Class Clostridia, although not all members of these Classes can form an endospore. While sporulating species are diverse (2, 9, 32), the fundamental process of forming an endospore appears the same (20, 22, 36). The sporulating cell divides asymmetrically resulting in two cells, a larger mother cell and a smaller forespore. The forespore is engulfed by the mother cell and prepares for dormancy by altering its cytoplasmic chemistry, obtaining a proteinaceous coat and a modified peptidoglycan called the cortex, and producing a variety of spore-specific proteins responsible for its resistance properties. Once the endospore has fully matured, it is released into the environment during mother-cell lysis. Due to these similarities, it is widely assumed that endospore-formers evolved from a common spore-forming ancestor. Certain bacteria from lineages outside the Firmicutes, like the Proteobacteria (*Myxococcus*), the Actinobacteria (*Streptomyces*), and the Cyanobacteria (*Anabaena* and *Nostoc*), have the ability to form spores or spore-like cells when nutrients become limited (1, 2, 18, 21, 31). While these developmental processes differ from endospore formation, all spore-formers create a dispersible, dormant cell in response to unfavorable environmental conditions, indicating a potential for a common evolutionary foundation behind spore formation in these organisms and the Firmicutes.

The underlying mechanisms and genetics responsible for the formation of a mature endospore have been extensively studied in the Gram-positive model organism *Bacillus subtilis* (13, 17, 49). Recent studies examining the regulons of the four sporulation-specific sigma factors (σ^F, σ^E, σ^G, and σ^K) and Spo0A have identified more than 700 genes that likely play a role during sporulation (11, 12, 30, 46, 50), making it an extremely complex process. *Bacillus*

subtilis employs an intercompartmental signaling network to ensure proper timing and localization of sigma factor activation and gene expression (17, 49). Similar regulatory systems are used by other endospore-formers (9, 33), but these systems are not nearly as well characterized.

Although endospore formation has been studied for over 50 years, few studies have examined the genetics behind the process outside the Bacilli until the past decade. While the *B. subtilis* model is a great basis for understanding endospore formation, there are stark differences in the reasons to sporulate and the mechanisms driving spore formation in other Firmicutes. For example, the guinea pig intestinal symbiont *Metabacterium polyspora* uses the formation of multiple endospores as a reproductive strategy in addition to binary fission, and spore formation appears to be a vital part of the *M. polyspora* symbiotic life cycle (5). *Clostridium acetobutylicum* forms an endospore to escape toxic byproducts produced during fermentation rather than in response to nutrient deprivation (34). Clostridia lack the phosphorelay system used by *B. subtilis* to integrate the environmental and cellular signals that launch endospore formation (47, 48). In addition, many Clostridia lack key genes involved in the activation of σ^K indicating a potential for lineage specific activation mechanisms (9, 33).

To fully understand the evolution of endospore formation and the fundamental characteristics driving it, the traits present in all endospore-formers should be identified. These characteristics would likely represent those of the common spore-forming ancestor. A study by Onyenwoke *et al.* (2004) examined the presence of 65 sporulation gene homologs from the *B. subtilis* model in endospore-forming and non-endospore-forming Firmicutes. They found that the genomes of species that have never been observed to form endospores, such as *Streptococcus* spp. and *Listeria* spp., still contained sequences similar to sporulation genes, including but not

limited to *spo0A*, *spoVB*, *spoVG* and *spoVK*. However, with the exception of *spoVG*, most of the scores from their BLAST analysis for these genes were too low for a definitive classification. Also, the authors conclude that spore-formers can have significantly different sporulation genes compared to those found in *B. subtilis*, and they list 13 genes found in some *Bacillus* spp. but not in any of the *Clostridium* spp. they examined. In a study by de Hoon *et al.* (2010), the genomes of 24 endospore-forming bacteria were examined for the presence of sporulation gene homologs. The researchers concluded that much of the core transcriptional regulatory network is conserved in the endospore-formers they examined. These studies laid the groundwork for examining evolution of endospore formation and its core developmental and regulatory mechanisms by examining the conservation of sporulation genes in the Firmicutes.

In this study, we expand the work from de Hoon *et al.* (2010) and Onyenwoke *et al.* (2004) by searching for sporulation gene homologs in 145 fully sequenced genomes from both endospore-forming and non-endospore forming Firmicutes. Based on a BLAST analysis to determine the distribution of *B. subtilis* sporulation genes in other endospore-forming bacteria, we were able to reveal a subset of genes present in all known endospore-formers. In addition, we identified 19 different species that have similar sporulation gene profiles to endospore-formers but have never been observed to form endospores, indicating either a potential to form endospores under the correct conditions or that they possess a defective sporulation program where restoring only a few genes would lead to endospore formation.

MATERIALS AND METHODS

Genomes used in this study. All fully sequenced genomes from the Firmicutes as of March 2011 were used in this study (Table A.2.1). In the instance where multiple genomes from the same species were available, a representative genome was chosen for comparison. The genomes of *Anabaena variabilis* ATCC 29413, *Nostoc punctiforme* PCC 73102, *Myxococcus xanthus* DK 1622, and *Streptomyces coelicolor* A3(2) were also used.

16S rRNA gene phylogenetic tree construction. Alignments of 16S rRNA gene sequences from the genomes of members of the Firmicutes were performed using Clustal X2 (23). Phylogenetic analysis was performed using the Phylip package v3.69 (14) by neighbor-joining after distance matrix construction using DNADIST. The 16S rRNA gene from *Veillonella parvula* was used as the outgroup.

Sporulation gene list construction. A full list of sporulation genes from *B. subtilis* was constructed using microarray studies examining the regulons of the four sporulation-specific sigma factors and Spo0A (11, 30, 46, 50). Other genes with sporulation functions listed in GenoList were also added (24). Genes with tested vegetative growth functions were removed from the list.

A smaller list of core sporulation genes was constructed as detailed in Chapter 2. Briefly, the initial list of sporulation genes was examined for conservation within the Bacilli and Clostridia. Genes that were not represented in 15 endospore-forming strains from each class were removed. Also, genes that were present in multiple non-sporulating species from the genera *Staphylococcus, Streptococcus, Lactobacillus*, *Listeria,* and/or *Enterococcus* were removed.

Comparative genomics. All the above genomes were searched by BLASTp (http://blast.ncbi.nlm.nih.gov/Blast.cgi) for sporulation proteins with cutoffs of 20% identity and 70% query coverage. This created a sporulation gene profile for each organism. Pearson's correlation was used to generate a distance matrix comparing the profiles. PHYLIP v3.69 was used to generate a neighbor-joining tree from the distance matrix. This analysis was repeated using only the genomes from endospore-forming Firmicutes, *A. variabilis, N. punctiforme, M. xanthus, S. coelicolor* A3(2), and the potential endospore-formers identified from the previous analysis with the large sporulation gene list. The potential spore-formers include *Acetohalobium arabaticum* DSM 5501, *Syntrophomonas wolfei* ssp. *wolfei* str. Goettingen, *Syntrophothermus lipocalidus* DSM 12680, *Symbiobacterium thermophilum* IAM, *Thermosediminibacter oceani* DSM16646, *Thermincola potens* JR, *Ammonifex degensii* KC4, *Natranaerobius thermophilus* JW/NM-WN-LF, the seven *Caldicellulosiruptor* spp., *Eubacterium rectale* ATCC 33656, *Eubacterium eligens* ATCC 27750, *Ruminococcus albus* 7, and *Ethanoligenens harbinense* YUAN 3.

RESULTS AND DISCUSSION

We collected the predicted protein sequences for genes identified as part of the Spo0A and the four sporulation-specific sigma factor regulons in *B. subtilis* (11, 30, 46, 50). When proteins with experimentally demonstrated vegetative growth functions were eliminated, the final list included 568 proteins. BLASTp was used to search for genes coding for these proteins in all completed genomes within the Firmicutes as well as the genomes of developmental models that produce other types of spores: *A. variabilis* and *N. punctiforme* (akinetes), *M. xanthus* (myxospores), and *S. coelicolor* (arthrospores). From these results, we generated a sporulation gene profile for each organism (Figure 1.1). As predicted, few sporulation genes were conserved in spore-formers and in non-spore-formers. Of the genes that were highly conserved, almost all function in metabolism (e.g. *gdh* and *glnQ*), peptidoglycan synthesis (e.g. *pbpA* and *pbpF*) or transport (e.g. *skfE* and *yheI*). Additionally, a few genes whose functions have yet to be determined (e.g. *yknU* and *yfnG*) were conserved in non-spore-formers. The broad distribution of these genes suggest a primary function during vegetative growth and that these genes have been co-opted for endospore formation. The sporulation-specific modulation of expression of genes that function during normal growth has been described before (49). For example, the cell division protein FtsZ is up-regulated to support asymmetric division during endospore formation (6, 35). Also, it is likely that at least some of the BLAST hits for these genes in non-spore-former genomes are due to the high similarity of other members from the same family of proteins that perform similar functions during vegetative growth. Similar to the findings of Onyenwoke *et al.* (2004), we found that other genes known to function during endospore formation but without an obvious vegetative growth function appear sporadically conserved in the non-endospore-formers.

Figure 1.1. The sporulation gene profiles of all fully sequenced Firmicutes. Heat map was generated using the percent identity for each homolog from the BLAST analysis. Names of the bacteria analyzed are on the left and the genes analyzed are across the top. White indicates a zero value, or no homolog present in that genome. Levels of grey indicate the varying percent identities. Black indicates an 80% or higher identity.

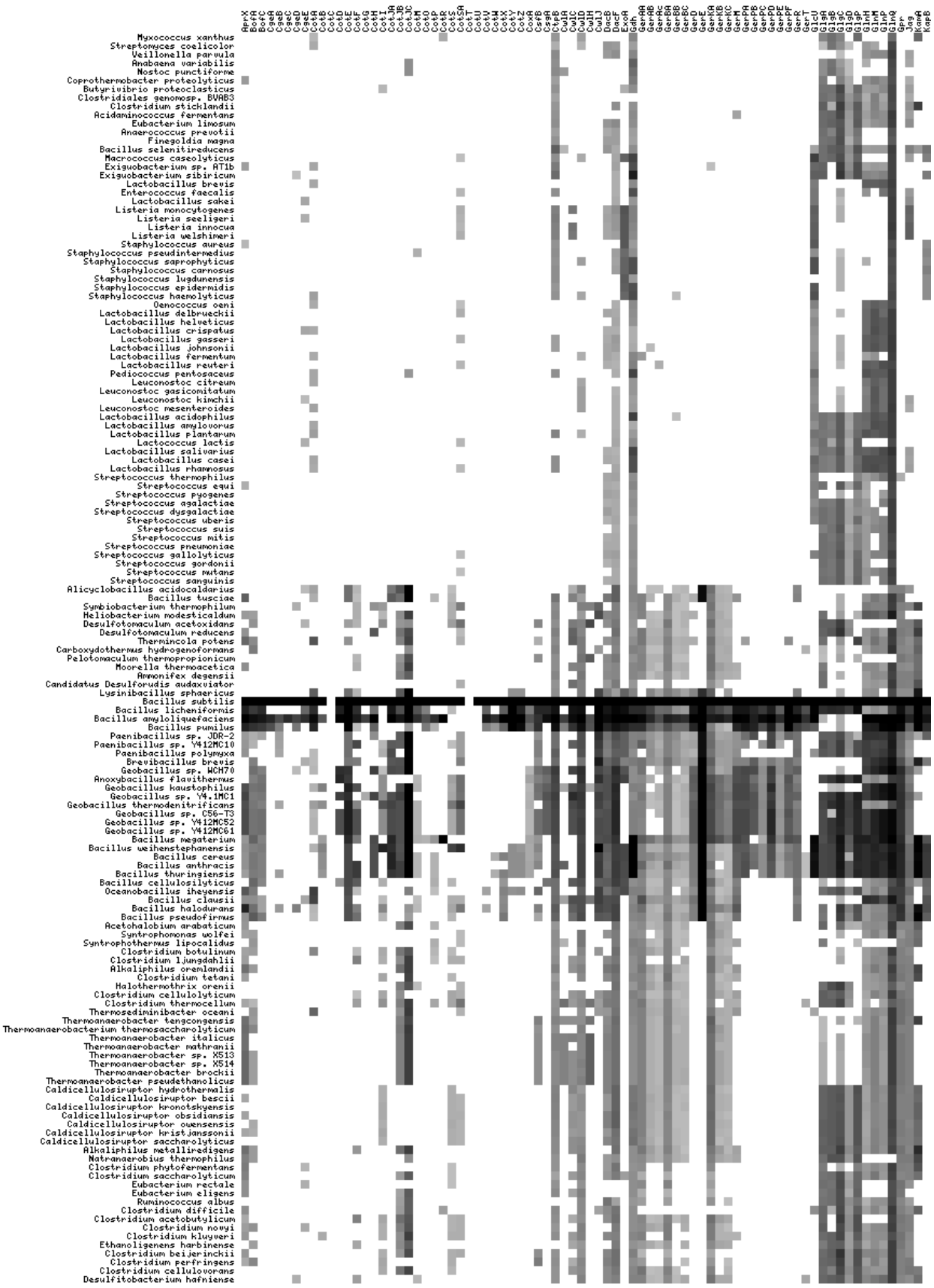

ApxX
BofA
BofC
CgeA
CgeB
CgeC
CgeD
CgeE
CgeF
CotB
CotC
CotD
CotE
CotG
CotH
CotI
CotJA
CotJB
CotJC
CotM
CotO
CotP
CotR
CotS
CotSA
CotT
CotU
CotV
CotW
CotX
CotY
CotZ
CoxA
CsfB
CsfA
CtpB
CwlA
CwlC
CwlD
CwlH
CwlJ
DacB
DacF
ExoA
GdhA
GerAA
GerAB
GerAC
GerBA
GerBB
GerBC
GerKA
GerE
GerKA
GerKB
GerKC
GerM
GerPA
GerPB
GerPC
GerPD
GerPE
GerPF
GerR
GerT
GlcU
GlnA
GlnB
GlnE
GlnQ
GlnP
GlnA
GlnH
GlnM
GlnP
GlnQ
Gpr
Jag
KamA
KapB
Myxococcus xanthus
Streptomyces coelicolor
Veillonella parvula
Anabaena variabilis
Nostoc punctiforme
Coprothermobacter proteolyticus
Butyrivibrio proteoclasticus
Clostridiales genomosp. BVAB3
Clostridium sticklandii
Acidaminococcus fermentans
Eubacterium limosum
Anaerococcus prevotii
Finegoldia magna
Bacillus selenitireducens
Macrococcus caseolyticus
Exiguobacterium sp. AT1b
Exiguobacterium sibiricum
Lactobacillus brevis
Enterococcus faecalis
Lactobacillus sakei
Listeria monocytogenes
Listeria seeligeri
Listeria innocua
Listeria welshimeri
Staphylococcus aureus
Staphylococcus pseudintermedius
Staphylococcus saprophyticus
Staphylococcus carnosus
Staphylococcus lugdunensis
Staphylococcus epidermidis
Staphylococcus haemolyticus
Oenococcus oeni
Lactobacillus delbrueckii
Lactobacillus helveticus
Lactobacillus crispatus
Lactobacillus gasseri
Lactobacillus johnsonii
Lactobacillus fermentum
Lactobacillus reuteri
Pediococcus pentosaceus
Leuconostoc citreum
Leuconostoc gasicomitatum
Leuconostoc kimchii
Leuconostoc mesenteroides
Lactobacillus acidophilus
Lactobacillus amylovorus
Lactobacillus plantarum
Lactococcus lactis
Lactobacillus salivarius
Lactobacillus casei
Lactobacillus rhamnosus
Streptococcus thermophilus
Streptococcus equi
Streptococcus pyogenes
Streptococcus agalactiae
Streptococcus dysgalactiae
Streptococcus uberis
Streptococcus suis
Streptococcus mitis
Streptococcus pneumoniae
Streptococcus gallolyticus
Streptococcus gordonii
Streptococcus mutans
Streptococcus sanguinis
Alicyclobacillus acidocaldarius
Bacillus tusciae
Symbiobacterium thermophilum
Heliobacterium modesticaldum
Desulfotomaculum acetoxidans
Desulfotomaculum reducens
Thermincola potens
Carboxydothermus hydrogenoformans
Pelotomaculum thermopropionicum
Moorella thermoacetica
Ammonifex degensii
Candidatus Desulforudis audaxviator
Lysinibacillus sphaericus
Bacillus subtilis
Bacillus licheniformis
Bacillus amyloliquefaciens
Bacillus pumilus
Paenibacillus sp. JDR-2
Paenibacillus sp. Y412MC10
Paenibacillus polymyxa
Brevibacillus brevis
Geobacillus sp. WCH70
Anoxybacillus flavithermus
Geobacillus kaustophilus
Geobacillus sp. Y4.1MC1
Geobacillus thermodenitrificans
Geobacillus sp. C56-T3
Geobacillus sp. Y412MC52
Geobacillus sp. Y412MC61
Bacillus megaterium
Bacillus weihenstephanensis
Bacillus cereus
Bacillus anthracis
Bacillus thuringiensis
Bacillus cellulosilyticus
Oceanobacillus iheyensis
Bacillus clausii
Bacillus halodurans
Bacillus pseudofirmus
Acetohalobium arabaticum
Syntrophomonas wolfei
Syntrophothermus lipocalidus
Clostridium botulinum
Clostridium ljungdahlii
Alkaliphilus oremlandii
Clostridium tetani
Halothermothrix orenii
Clostridium cellulolyticum
Clostridium thermocellum
Thermosediminibacter oceani
Thermoanaerobacter tengcongensis
Thermoanaerobacterium thermosaccharolyticum
Thermoanaerobacter italicus
Thermoanaerobacter mathranii
Thermoanaerobacter sp. X513
Thermoanaerobacter sp. X514
Thermoanaerobacter brockii
Thermoanaerobacter pseudethanolicus
Caldicellulosiruptor hydrothermalis
Caldicellulosiruptor bescii
Caldicellulosiruptor kronotskyensis
Caldicellulosiruptor obsidiansis
Caldicellulosiruptor owensensis
Caldicellulosiruptor kristjanssonii
Caldicellulosiruptor saccharolyticus
Alkaliphilus metalliredigens
Natranaerobius thermophilus
Clostridium phytofermentans
Clostridium saccharolyticum
Eubacterium rectale
Eubacterium eligens
Ruminococcus albus
Clostridium difficile
Clostridium acetobutylicum
Clostridium novyi
Clostridium kluyveri
Ethanoligenens harbinense
Clostridium beijerinckii
Clostridium perfringens
Clostridium cellulovorans
Desulfitobacterium hafniense

Figure 1.1 (Continued)

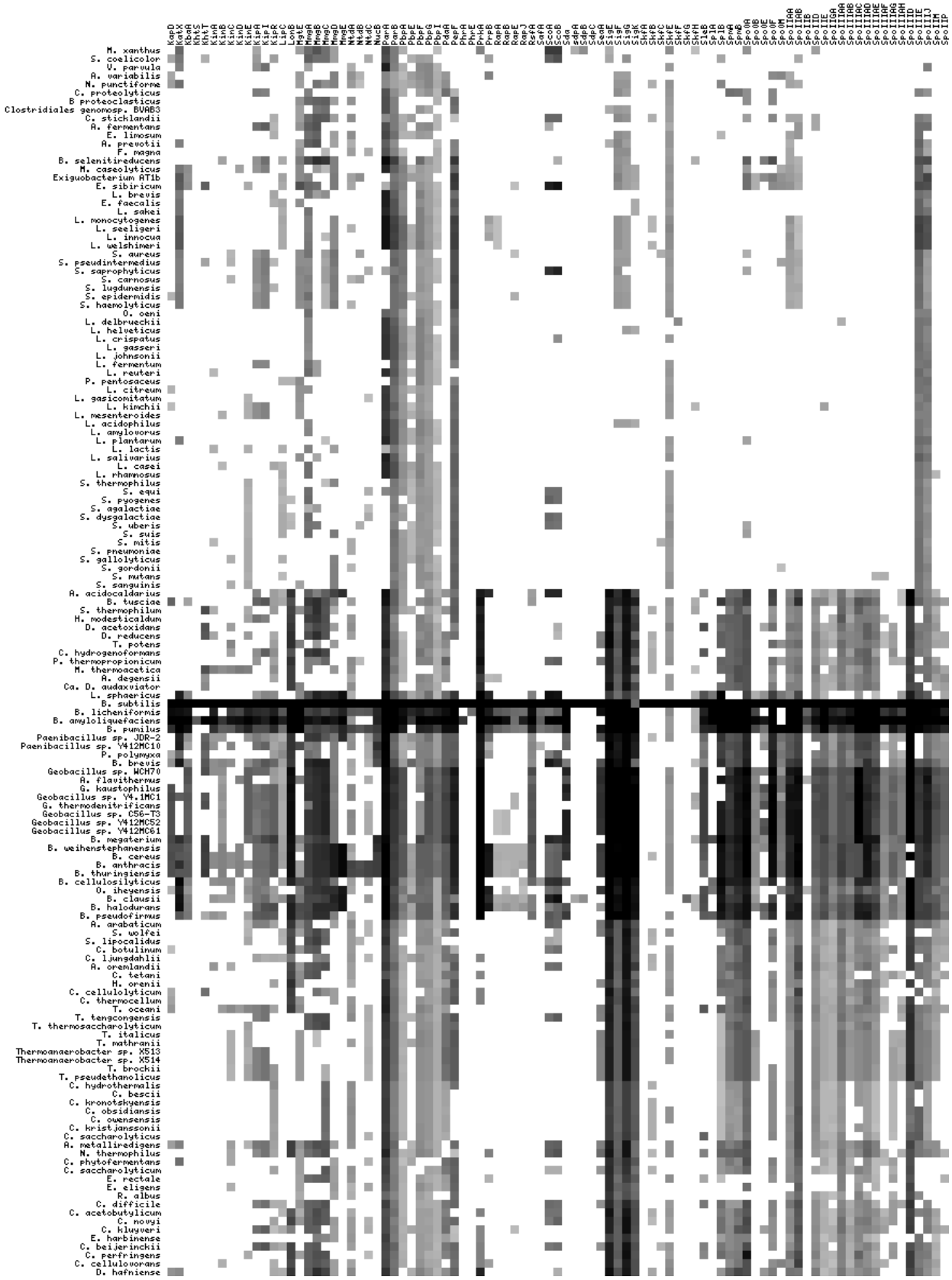


Figure 1.1 (Continued)

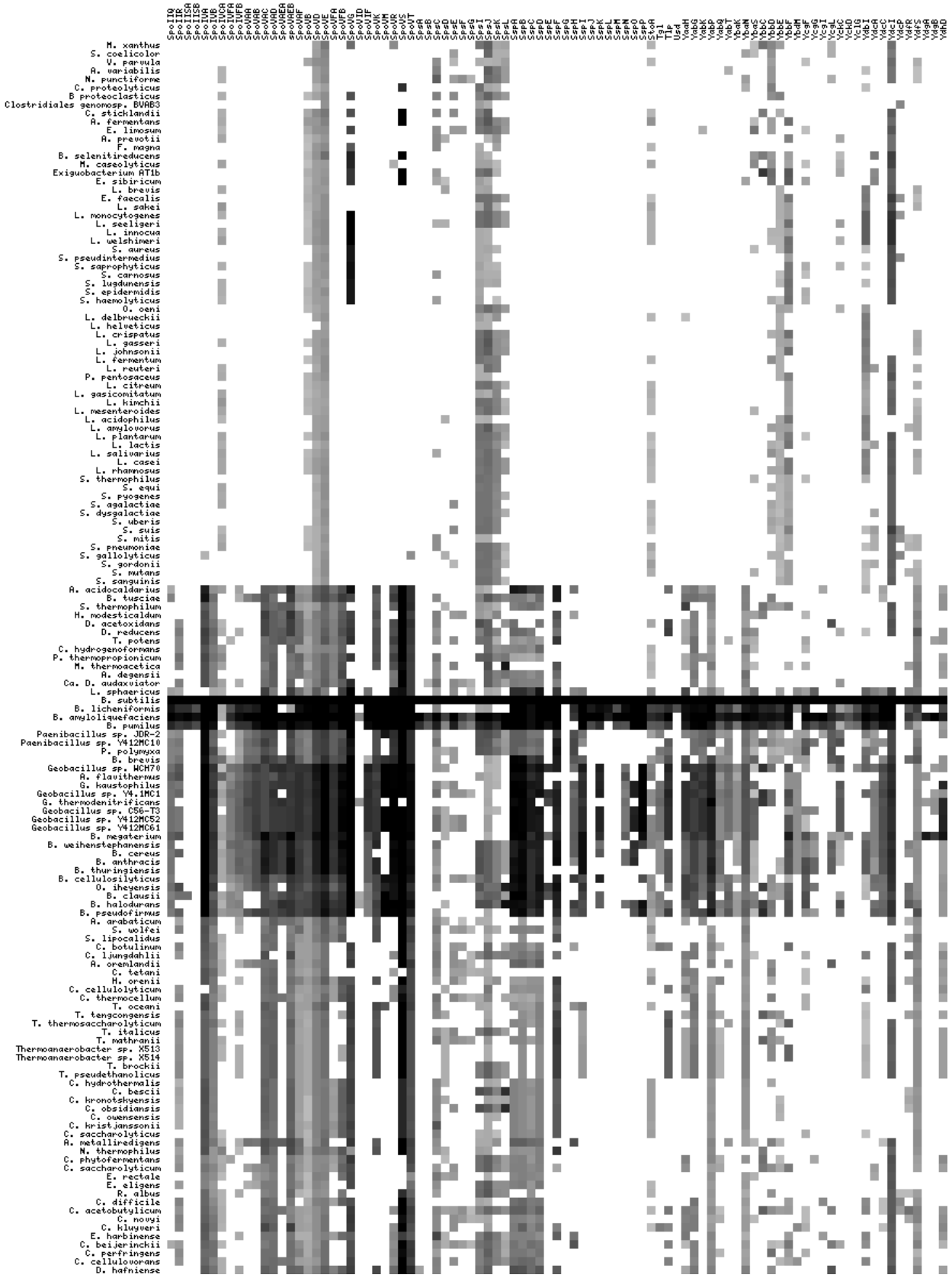


Figure 1.1 (Continued)

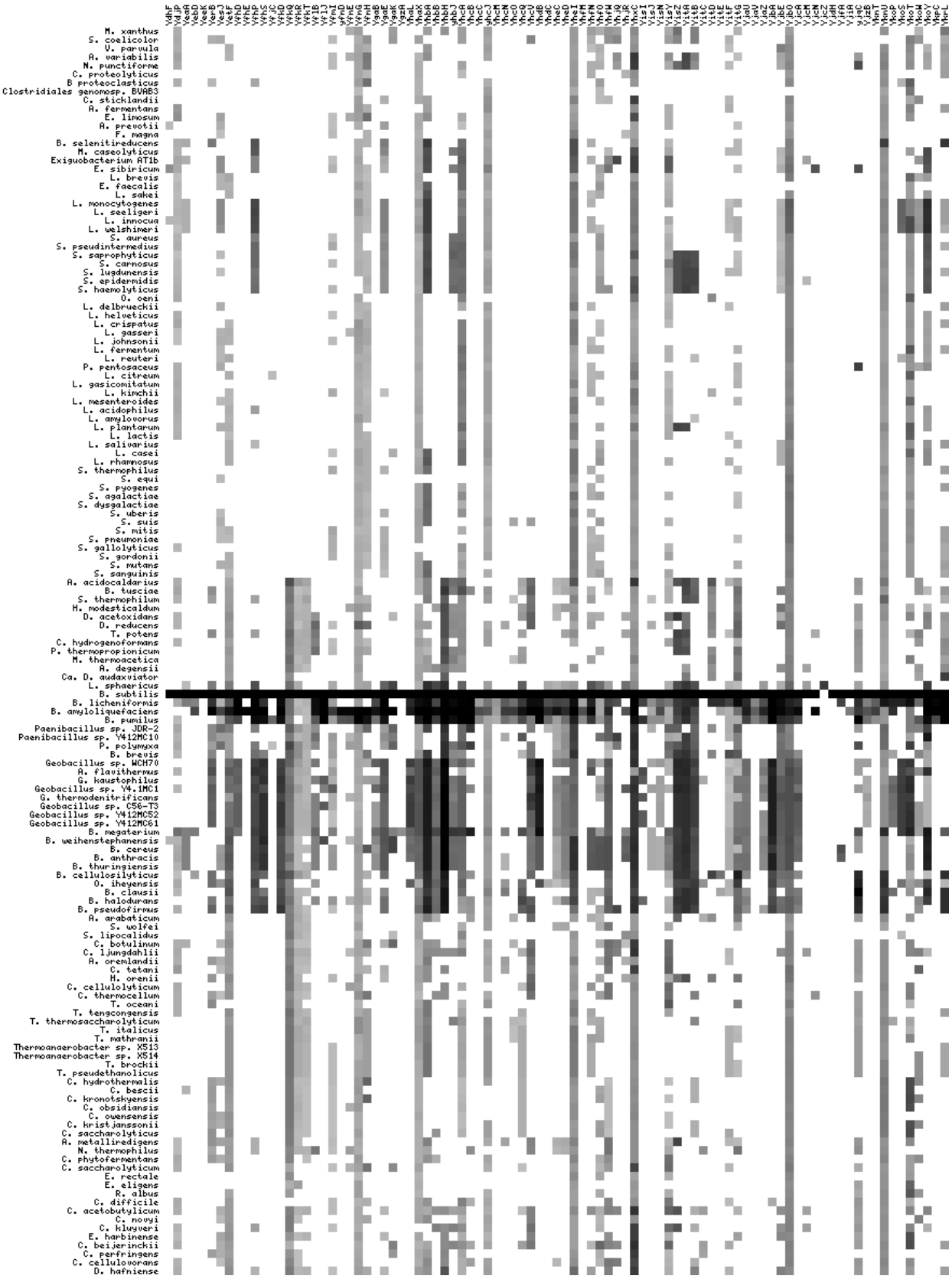


Figure 1.1 (Continued)

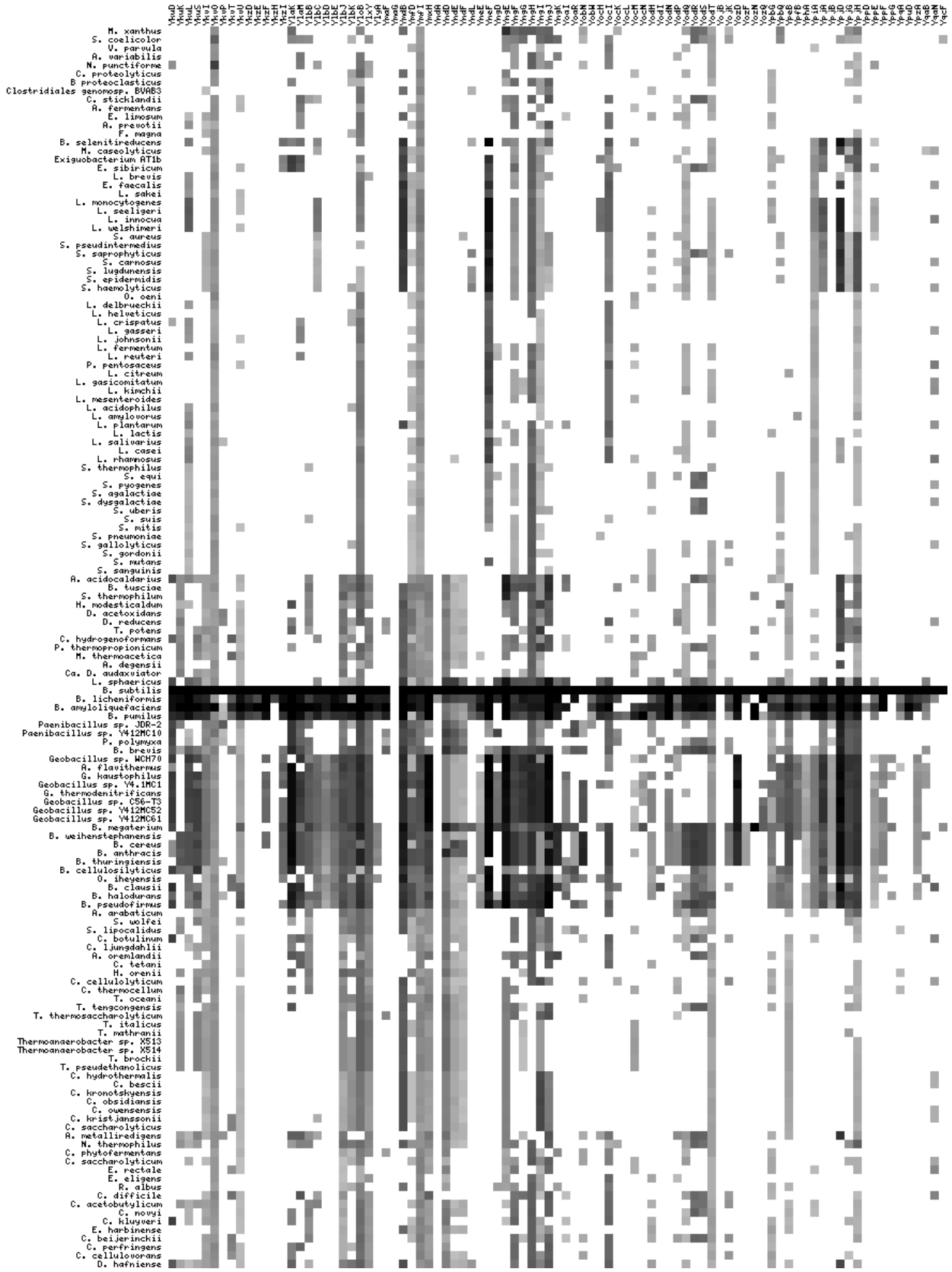


Figure 1.1 (Continued)

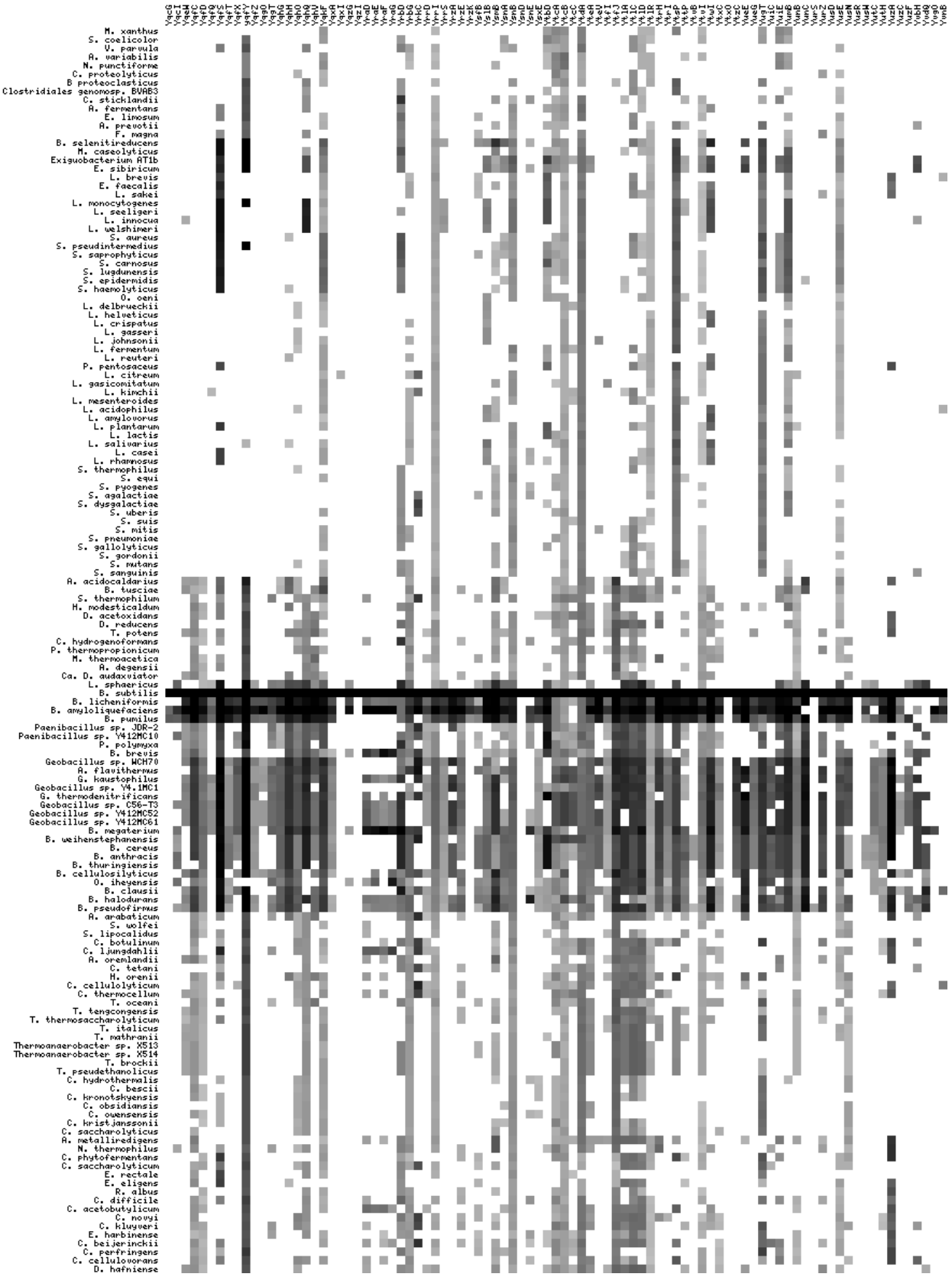


Figure 1.1 (Continued)

To take a more holistic approach and see how these overall profiles relate to one another, the sporulation profiles were compared using a Pearson's coefficient analysis (Figure 1.2). Two distinct clusters of endospore-formers and non-endospore-formers appeared, and within each of these groups, bacteria from related lineages (based on 16S rRNA gene trees) (Figure 1.3) grouped together. It is likely that the clustering pattern within the endospore-forming group represents lineage-specific divergences in the evolution of endospore formation.

Strangely, *Clostridium sticklandii* grouped with the non-endospore-former clade. When it was first described *C. sticklandii* was observed to form endospores and it has been used as an important model in studying the breakdown of amino acids by the Stickland reaction (15, 45). To the best of our knowledge, there have been no studies specifically focused on describing sporulation in this organism. Our results indicate that the specific strain whose genome we used, *Clostridium sticklandii* DSM 519, does not have the genetic capacity to form an endospore as its sporulation gene profile more closely resembles that of a non-endospore-former rather than any other bacterium in the genus *Clostridium*. Specifically, *C. sticklandii* does not contain homologs for two (*sigE* and *sigK*) of the four sporulation-specific sigma factors, which are present in all endospore-formers examined, and it seems unlikely a different mechanism for transcriptional regulation has evolved only in *C. sticklandii*. It is also missing key components of the intercompartmental signaling pathway regulating the activition of the sigma factors (*spoIIE*, *bofA*, *spoIIGA*, *spoIIR*, the *spoIIIA* operon and *spoIVB*), genes important for engulfment (*spoIIM* and *spoIIP*) and genes whose products are responsible for making the spore resistant (the *spoVA* operon). We were unable to identify *bofA*, *spoIIGA*, *spoIIR*, *spoIIM* and *spoIIP* in the genome of the closest spore-forming relative of *C. sticklandii*, *Clostridium difficile*, but it contains the other genes listed above and many more peripheral ones that *C. sticklandii* does not. The similar

Figure 1.2. Tree comparing the sporulation gene profiles of all fully sequenced Firmicutes.

This tree was generated using the UPGMA method with the PHYLIP 3.65 package.

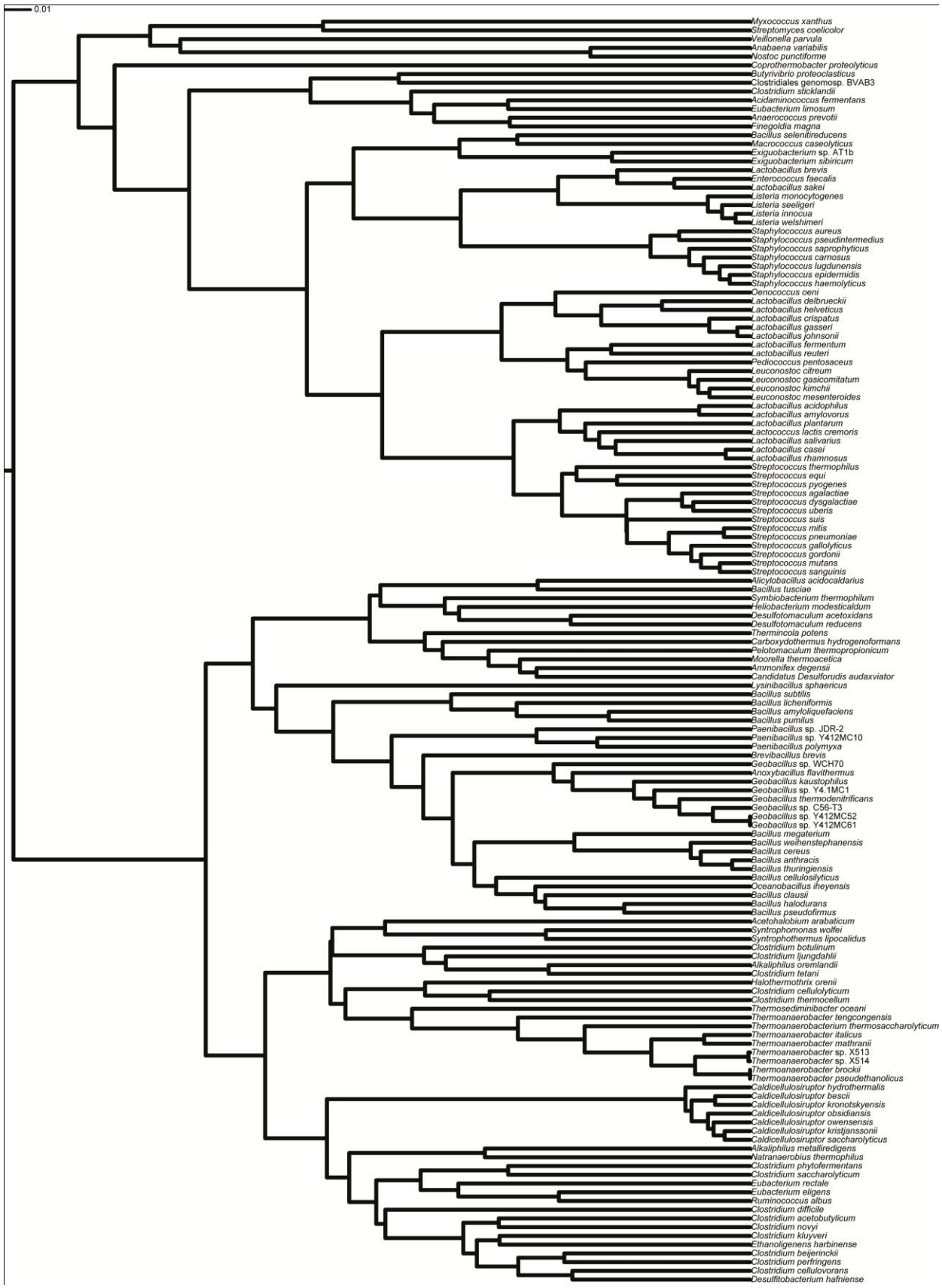
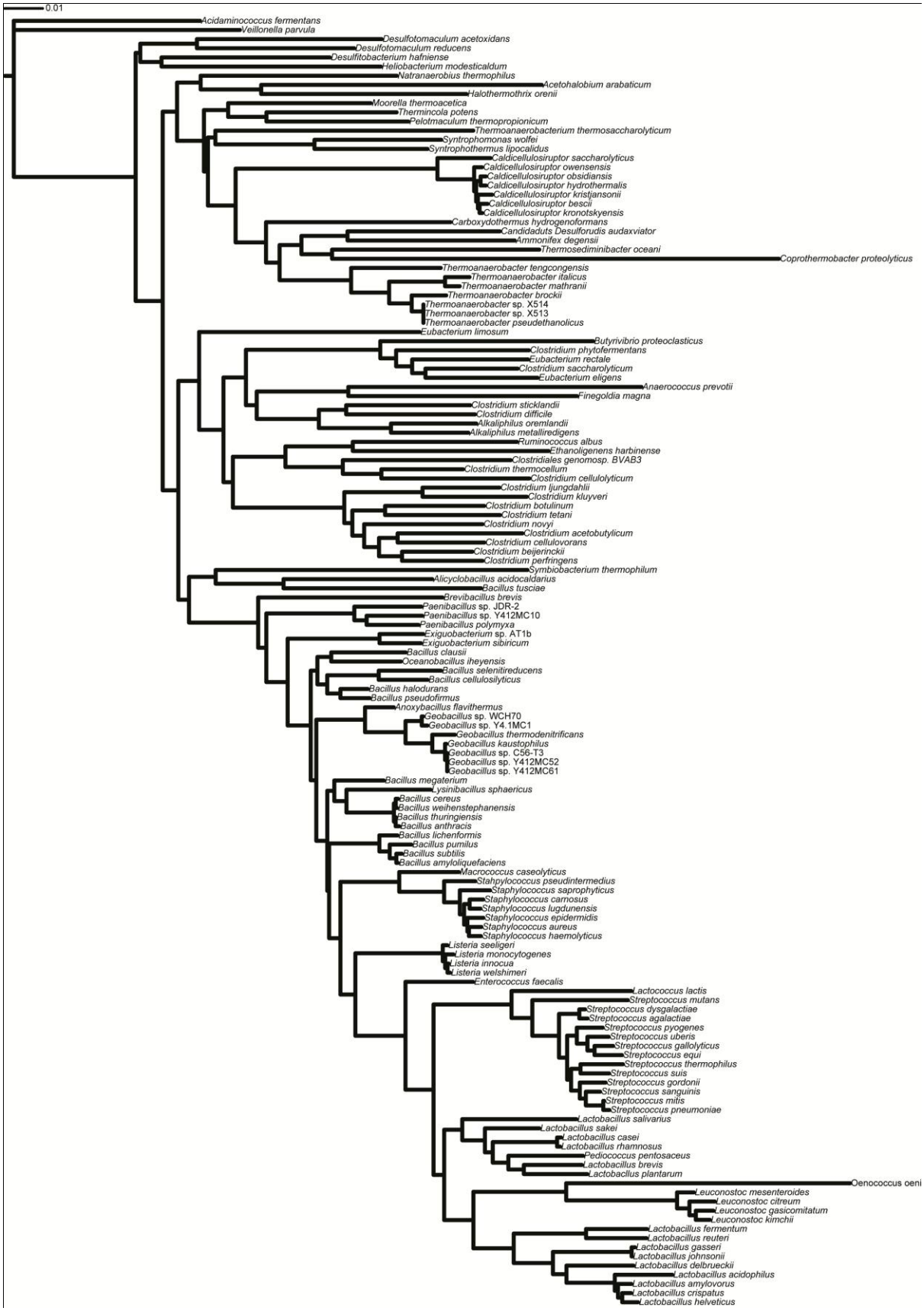

0.01
Myxococcus xanthus
Streptomyces coelicolor
Veillonella parvula
Anabaena variabilis
Nostoc punctiforme
Coprothermobacter proteolyticus
Butyrivibrio proteoclasticus
Clostridiales genomosp. BVAB3
Clostridium sticklandii
Acidaminococcus fermentans
Eubacterium limosum
Anaerococcus prevotii
Finegoldia magna
Bacillus selenitireducens
Macrococcus caseolyticus
Exiguobacterium sp. AT1b
Exiguobacterium sibiricum
Lactobacillus brevis
Enterococcus faecalis
Lactobacillus sakei
Listeria monocytogenes
Listeria seeligeri
Listeria innocua
Listeria welshimeri
Staphylococcus aureus
Staphylococcus pseudintermedius
Staphylococcus saprophyticus
Staphylococcus carnosus
Staphylococcus lugdunensis
Staphylococcus epidermidis
Staphylococcus haemolyticus
Oenococcus oeni
Lactobacillus delbrueckii
Lactobacillus helveticus
Lactobacillus crispatus
Lactobacillus gasseri
Lactobacillus johnsonii
Lactobacillus fermentum
Lactobacillus reuteri
Pediococcus pentosaceus
Leuconostoc citreum
Leuconostoc gasicomitatum
Leuconostoc kimchii
Leuconostoc mesenteroides
Lactobacillus acidophilus
Lactobacillus amylovorus
Lactobacillus plantarum
Lactococcus lactis cremoris
Lactobacillus salivarius
Lactobacillus casei
Lactobacillus rhamnosus
Streptococcus thermophilus
Streptococcus equi
Streptococcus pyogenes
Streptococcus agalactiae
Streptococcus dysgalactiae
Streptococcus uberis
Streptococcus suis
Streptococcus mitis
Streptococcus pneumoniae
Streptococcus gallolyticus
Streptococcus gordonii
Streptococcus mutans
Streptococcus sanguinis
Alicyclobacillus acidocaldarius
Bacillus tusciae
Symbiobacterium thermophilum
Heliobacterium modesticaldum
Desulfotomaculum acetoxidans
Desulfotomaculum reducens
Thermincola potens
Carboxydothermus hydrogenoformans
Pelotomaculum thermopropionicum
Moorella thermoacetica
Ammonifex degensii
Candidatus Desulforudis audaxviator
Lysinibacillus sphaericus
Bacillus subtilis
Bacillus licheniformis
Bacillus amyloliquefaciens
Bacillus pumilus
Paenibacillus sp. JDR-2
Paenibacillus sp. Y412MC10
Paenibacillus polymyxa
Brevibacillus brevis
Geobacillus sp. WCH70
Anoxybacillus flavithermus
Geobacillus kaustophilus
Geobacillus sp. Y4.1MC1
Geobacillus thermodenitrificans
Geobacillus sp. C56-T3
Geobacillus sp. Y412MC52
Geobacillus sp. Y412MC61
Bacillus megaterium
Bacillus weihenstephanensis
Bacillus cereus
Bacillus anthracis
Bacillus thuringiensis
Bacillus cellulosilyticus
Oceanobacillus iheyensis
Bacillus clausii
Bacillus halodurans
Bacillus pseudofirmus
Acetohalobium arabaticum
Syntrophomonas wolfei
Syntrophothermus lipocalidus
Clostridium botulinum
Clostridium ljungdahlii
Alkaliphilus oremlandii
Clostridium tetani
Halothermothrix orenii
Clostridium cellulolyticum
Clostridium thermocellum
Thermosediminibacter oceani
Thermoanaerobacter tengcongensis
Thermoanaerobacterium thermosaccharolyticum
Thermoanaerobacter italicus
Thermoanaerobacter mathranii
Thermoanaerobacter sp. X513
Thermoanaerobacter sp. X514
Thermoanaerobacter brockii
Thermoanaerobacter pseudethanolicus
Caldicellulosiruptor hydrothermalis
Caldicellulosiruptor bescii
Caldicellulosiruptor kronotskyensis
Caldicellulosiruptor obsidiansis
Caldicellulosiruptor owensensis
Caldicellulosiruptor kristjanssonii
Caldicellulosiruptor saccharolyticus
Alkaliphilus metalliredigens
Natranaerobius thermophilus
Clostridium phytofermentans
Clostridium saccharolyticum
Eubacterium rectale
Eubacterium eligens
Ruminococcus albus
Clostridium difficile
Clostridium acetobutylicum
Clostridium novyi
Clostridium kluyveri
Ethanoligenens harbinense
Clostridium beijerinckii
Clostridium perfringens
Clostridium cellulovorans
Desulfitobacterium hafniense

Figure 1.3. 16S Phylogenetic tree of the Firmicutes. Neighbor-joining tree based on 16S rRNA gene sequences. The tree was rooted with the 16S rRNA gene sequence from *Veillonella parvula*.


0.01
Acidaminococcus fermentans
Veillonella parvula
Desulfotomaculum acetoxidans
Desulfotomaculum reducens
Desulfitobacterium hafniense
Heliobacterium modesticaldum
Natranaerobius thermophilus
Acetohalobium arabaticum
Halothermothrix orenii
Moorella thermoacetica
Thermincola potens
Pelotmaculum thermopropionicum
Thermoanaerobacterium thermosaccharolyticum
Syntrophomonas wolfei
Syntrophothermus lipocalidus
Caldicellulosiruptor saccharolyticus
Caldicellulosiruptor owensensis
Caldicellulosiruptor obsidiansis
Caldicellulosiruptor hydrothermalis
Caldicellulosiruptor kristjansonii
Caldicellulosiruptor bescii
Caldicellulosiruptor kronotskyensis
Carboxydothermus hydrogenoformans
Candidaduts Desulforudis audaxviator
Ammonifex degensii
Thermosediminibacter oceani
Coprothermobacter proteolyticus
Thermoanaerobacter tengcongensis
Thermoanaerobacter italicus
Thermoanaerobacter mathranii
Thermoanaerobacter brockii
Thermoanaerobacter sp. X514
Thermoanaerobacter sp. X513
Thermoanaerobacter pseudethanolicus
Eubacterium limosum
Butyrivibrio proteoclasticus
Clostridium phytofermentans
Eubacterium rectale
Clostridium saccharolyticum
Eubacterium eligens
Anaerococcus prevotii
Finegoldia magna
Clostridium sticklandii
Clostridium difficile
Alkaliphilus oremlandii
Alkaliphilus metalliredigens
Ruminococcus albus
Ethanoligenens harbinense
Clostridiales genomosp. BVAB3
Clostridium thermocellum
Clostridium cellulolyticum
Clostridium ljungdahlii
Clostridium kluyveri
Clostridium botulinum
Clostridium tetani
Clostridium novyi
Clostridium acetobutylicum
Clostridium cellulovorans
Clostridium beijerinckii
Clostridium perfringens
Symbiobacterium thermophilum
Alicyclobacillus acidocaldarius
Bacillus tusciae
Brevibacillus brevis
Paenibacillus sp. JDR-2
Paenibacillus sp. Y412MC10
Paenibacillus polymyxa
Exiguobacterium sp. AT1b
Exiguobacterium sibiricum
Bacillus clausii
Oceanobacillus iheyensis
Bacillus selenitireducens
Bacillus cellulosilyticus
Bacillus halodurans
Bacillus pseudofirmus
Anoxybacillus flavithermus
Geobacillus sp. WCH70
Geobacillus sp. Y4.1MC1
Geobacillus thermodenitrificans
Geobacillus kaustophilus
Geobacillus sp. C56-T3
Geobacillus sp. Y412MC52
Geobacillus sp. Y412MC61
Bacillus megaterium
Lysinibacillus sphaericus
Bacillus cereus
Bacillus weihenstephanensis
Bacillus thuringiensis
Bacillus anthracis
Bacillus lichenformis
Bacillus pumilus
Bacillus subtilis
Bacillus amyloliquefaciens
Macrococcus caseolyticus
Stahylococcus pseudintermedius
Staphylococcus saprophyticus
Staphylococcus carnosus
Staphylococcus lugdunensis
Staphylococcus epidermidis
Staphylococcus aureus
Staphylococcus haemolyticus
Listeria seeligeri
Listeria monocytogenes
Listeria innocua
Listeria welshimeri
Enterococcus faecalis
Lactococcus lactis
Streptococcus mutans
Streptococcus dysgalactiae
Streptococcus agalactiae
Streptococcus pyogenes
Streptococcus uberis
Streptococcus gallolyticus
Streptococcus equi
Streptococcus thermophilus
Streptococcus suis
Streptococcus gordonii
Streptococcus sanguinis
Streptococcus mitis
Streptococcus pneumoniae
Lactobacillus salivarius
Lactobacillus sakei
Lactobacillus casei
Lactobacillus rhamnosus
Pediococcus pentosaceus
Lactobacillus brevis
Lactobacillus plantarum
Oenococcus oeni
Leuconostoc mesenteroides
Leuconostoc citreum
Leuconostoc gasicomitatum
Leuconostoc kimchii
Lactobacillus fermentum
Lactobacillus reuteri
Lactobacillus gasseri
Lactobacillus johnsonii
Lactobacillus delbrueckii
Lactobacillus acidophilus
Lactobacillus amylovorus
Lactobacillus crispatus
Lactobacillus helveticus

absence of highly conserved sporulation genes in *C. sticklandii* and *C. difficile* suggests that this lineage diverged from the *B. subtilis* model of endospore formation early on and may employ a different set genes serving these functions.

Within the endospore-former clade, 19 bacteria that have been characterized as non-sporulating species are represented: *Symbiobacterium thermophilum, Thermincola potens, Ammonifex degensii, Acetohalobium arabaticum, Syntrophomonas wolfei, Syntrophothermus lipocalidus, Thermosediminibacter oceani, Caldicellulosiruptor hydrothermalis, Caldicellulosiruptor bescii, Caldicellulosiruptor kronitskyensis, Caldicellulosiruptor obsidiansis, Caldicellulosiruptor owensensis, Caldicellulosiruptor kristjanssonii, Caldicellulosiruptor saccharolyticus, Eubacterium rectale, Eubacterium eligens, Natranaerobius thermophilus Ruminococcus albus* and *Ethanoligenens harbinense*. Our analysis suggests that these organisms could produce endospores under the correct conditions, or they have lost this ability recently in their evolutionary histories. From this point on, these bacteria will be referred to as potential endospore-formers.

To further examine the evolution of endospore formation, we searched for a core subset of sporulation genes in the genomes of all known endospore-formers and the potential endospore-formers (Table A.2.2). The development of this list of 147 genes is described in detail in Chapter 2. The list contains sporulation-specific genes conserved in both Class Bacilli and Class Clostridia. A core sporulation gene profile was generated for each organism (Figure 1.4). The profiles were compared by Pearson's coefficient analysis and a tree was constructed from the results (Figure 1.5). As above, bacteria from similar genera have highly similar profiles and group together. Of the core genes, 61 are conserved in greater than 90% of the known endospore-formers (Table 1.1). The conservation of these genes suggests they were part of an

Figure 1.4. The core sporulation gene profiles of endospore-formers and potential endospore-formers. Heat map was generated using the percent identity for each homolog from the BLAST analysis. Names of the bacteria analyzed are on the left and the genes analyzed are across the top. White indicates a zero value, or no homolog present in that genome. Levels of grey indicate the varying percent identities. Black indicates an 80% or higher identity.

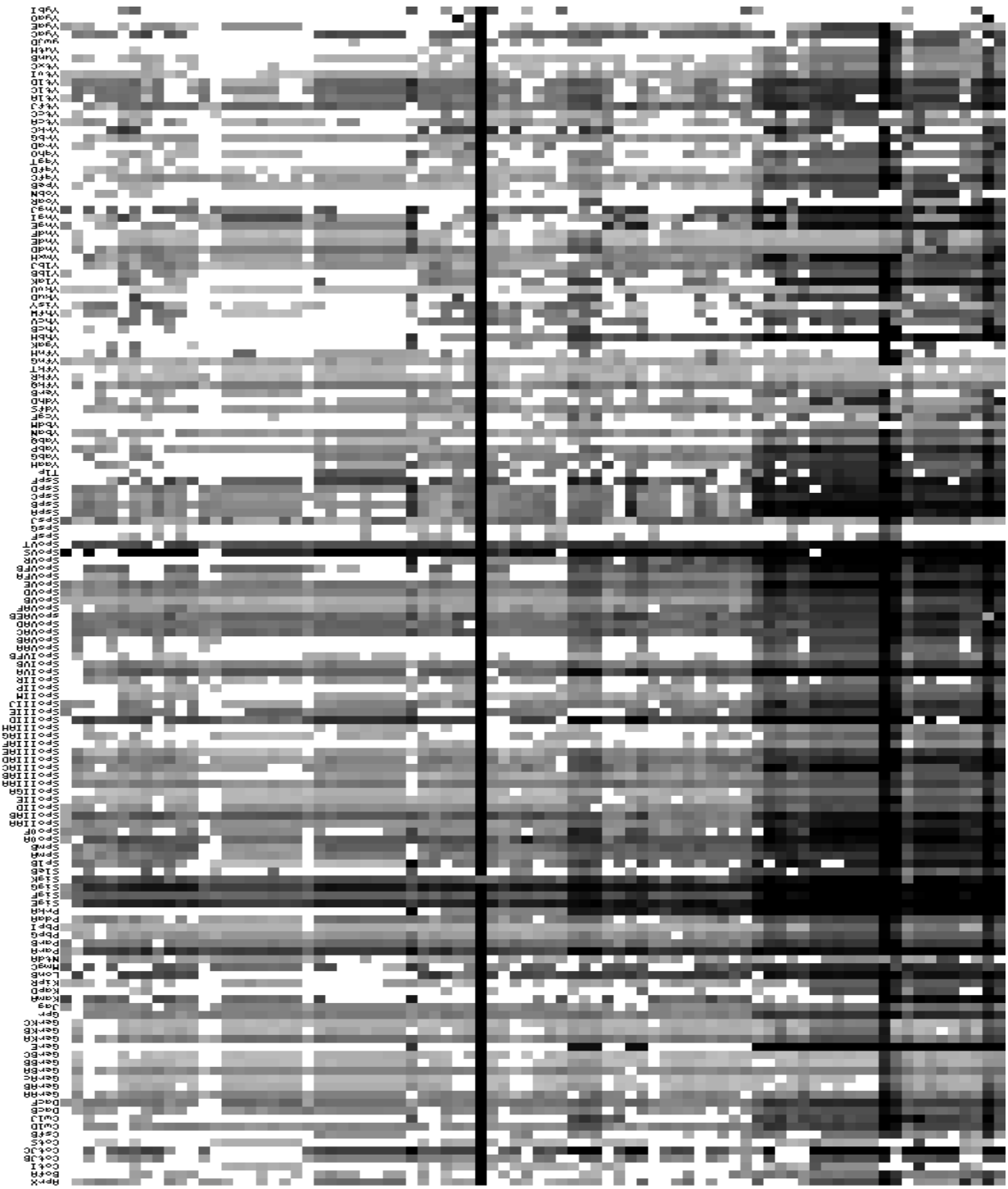

Clostridium stickl andii
Alkaliphilus metalliredigens
Clostridium difficile
Clostridium phytofermentans
Clostridium saccharolyticum
Acetohalobium arabaticum
Clostridium botulinum
Thermosediminibacter oceani
Alkaliphilus oremlandii
Natranaerobius thermophilus
Syntrophomonas wolfei
Syntrophothermus lipocalidus
Eubacterium eligens
Eubacterium rectale
Caldicellulosiruptor bescii
Caldicellulosiruptor kronotskyensis
Caldicellulosiruptor hydrothermalis
Caldicellulosiruptor hrist_jansonii
Caldicellulosiruptor saccharolyticus
Caldicellulosiruptor obsidiensis
Caldicellulosiruptor owensensis
Ruminococcus albus
Thermoanaerobacter tengcongensis
Thermoanaerobacterium thermosaccharolyticum
Thermoanaerobacter sp. X513
Thermoanaerobacter sp. X514
Thermoanaerobacter wiegelii
Thermoanaerobacter italicus
Thermoanaerobacter brockii
Thermoanaerobacter pseudethanolicus
Lysinibacillus sphaericus
Clostridium novyi
Halothermothrix orenii
Clostridium thermocellum
Clostridium kluyveri
Clostridium limosum
Bacillus subtilis
Clostridium cellulolyticum
Ethanoligenens harbinense
Clostridium beijerinckii
Clostridium perfringens
Clostridium cellulovorans
Clostridium acetobutylicum
Clostridium tetani
Paenibacillus polymyxa
Paenibacillus sp. JDR-2
Paenibacillus sp. Y412MC10
Symbiobacterium thermophilum
Heliobacterium modesticaldum
Alicyclobacillus acidocaldarius
Bacillus tusciae
Candidatus Desulforudis audaxviator
Ammonifex degensii
Moorella thermoacetica
Thermincola potens
Desulfitobacterium hafniense
Desulfotomaculum reducens
Pelotomaculum thermopropionicum
Carboxydothermus hydrogenoformans
Desulfotomaculum acetoxidans
Oceanobacillus iheyensis
Geobacillus sp. WCH70
Anoxybacillus flavithermus
Bacillus cellulosilyticus
Bacillus clausii
Geobacillus thermodenitrificans
Geobacillus kaustophilus
Geobacillus sp. Y4.1MC1
Geobacillus sp. C56-T3
Geobacillus sp. Y412MC52
Geobacillus sp. Y412MC61
Bacillus amyloliquefaciens
Bacillus pumilus
Brevibacillus brevis
Bacillus weihenstephanensis
Bacillus cereus
Bacillus anthracis
Bacillus thuringiensis
Bacillus halodurans
Bacillus pseudofirmus
Bacillus licheniformis
Bacillus megaterium

Figure 1.5. Tree comparing the core sporulation gene profiles of endospore-formers and potential endospore-formers. This tree was generated using the UPGMA method with the PHYLIP 3.65 package.


0.01
Clostridium sticklandii
Alkaliphilus metalliredigens
Clostridium difficile
Clostridium phytofermentans
Clostridium saccharolyticum
Acetohalobium arabaticum
Clostridium botulinum
Thermosediminibacter oceani
Alkaliphilus oremlandii
Natranaerobius thermophilus
Syntrophomonas wolfei
Syntrophothermus lipocalidus
Eubacterium eligens
Eubacterium rectale
Caldicellulosiruptor bescii
Caldicellulosiruptor kronotskyensis
Caldicellulosiruptor hydrothermalis
Caldicellulosiruptor kristjanssonii
Caldicellulosiruptor saccharolyticus
Caldicellulosiruptor obsidiansis
Caldicellulosiruptor owensensis
Ruminococcus albus
Thermoanaerobacter tengcongensis
Thermoanaerobacterium thermosaccharolyticum
Thermoanaerobacter sp. X513
Thermoanaerobacter sp. X514
Thermoanaerobacter mathranii
Thermoanaerobacter italicus
Thermoanaerobacter brockii
Thermoanaerobacter pseudethanolicus
Lysinibacillus sphaericus
Clostridium novyi
Halothermothrix orenii
Clostridium thermocellum
Clostridium kluyveri
Clostridium ljungdahlii
Bacillus subtilis
Clostridium cellulolyticum
Ethanoligenens harbinense
Clostridium beijerinckii
Clostridium perfringens
Clostridium cellulovorans
Clostridium acetobutylicum
Clostridium tetani
Paenibacillus polymyxa
Paenibacillus sp. JDR-2
Paenibacillus sp. Y412MC10
Symbiobacterium thermophilum
Heliobacterium modesticaldum
Alicyclobacillus acidocaldarium
Bacillus tusciae
Candidatus Desulforudis audaxviator
Ammonifex degensii
Moorella thermoacetica
Thermincola potens
Desulfitobacterium hafniense
Desulfotomaculum reducens
Pelotomaculum thermopropionicum
Carboxydothermus hydrogenoformans
Desulfotomaculum acetoxidans
Oceanobacillus iheyensis
Geobacillus sp. WCH70
Anoxybacillus flavithermus
Bacillus cellulosilyticus
Bacillus clausii
Geobacillus thermodenitrificans
Geobacillus kaustophilus
Geobacillus sp. Y4.1MC1
Geobacillus sp. C56-T3
Geobacillus sp. Y412MC52
Geobacillus sp. Y412MC61
Bacillus amyloliquefaciens
Bacillus pumilus
Bacillus brevis
Bacillus weihenstephanensis
Bacillus cereus
Bacillus anthracis
Bacillus thuringiensis
Bacillus halodurans
Bacillus pseudofirmus
Bacillus lichenformis
Bacillus megaterium

Table 1.1 Core genes conserved in 90% of endospore-formers

GENE	FUNCTION
cwlD	N-acetylmuramoyl-L-alanine amidase/germination
dacF	D-alanyl-D-alanine carboxypeptidase (penicilin binding protein)
gerAA	component of the germination receptor GerA
gerBA	component of germinant receptor B
gpr	germination protease
jag	SpoIIIJ-associated RNA/ssDNA-binding protein
pbpG	penicillin-binding protein (also known as ywhe)
pbpI	penicillin-binding protein PBP4B/mother cell specific
pdaA	exported N-acetylmuramic acid deacetylase/cortex lysis
sigE	sporulation sigma factor/mother cell only
sigF	sporulation sigma factor/forespore only
sigG	sporulation sigma factor/forespore only
sigK	sporulation sigma factor/mother cell only
soj	chromosome partitioning protein/transcriptional regulator/negative regulation of sporulation initiation
spmA	spore maturation protein/spore dehydration
spmB	spore maturation protein/spore dehydration
spo0A	two-component response regulator central for the initiation of sporulation/"master regulator"
spo0J	site-specific DNA-binding protein/chromosome positioning near the pole and transport through the polar septum /antagonist of
spoIIAA	anti-anti-sigma factor (antagonist of SpoIIAB)
spoIIAB	anti-σ^F factor
spoIID	autolysin required for complete dissolution of the asymmetric septum
spoIIE	serine phosphatase (σ^F activation) /polar septum formation
spoIIGA	protease processing pro-σ^E
spoIIIAA	ATP-binding stage III sporulation protein/mother cell signalling for activation of σ^G
spoIIIAB	stage III sporulation protein/mother cell signalling for activation of σ^G
spoIIIAC	stage III sporulation protein/mother cell signalling for activation of σ^G
spoIIIAD	stage III sporulation protein/mother cell signalling for activation of σ^G
spoIIIAE	stage III sporulation protein/mother cell signalling for activation of σ^G
spoIIID	transcriptional regulator of σ^E and σ^K-dependent genes
spoIIM	autolysin component for dissolution of the septal cell wall
spoIVA	morphogenetic protein required for proper spore cortex formation and coat assembly
spoIVB	regulatory membrane-associated serine protease/intercompartmental signalling of pro-σ^K processing and activation in the
spoVAC	essential for dipicolinic acid uptake by the developing spore
spoVAD	essential for dipicolinic acid uptake by the developing spore
spoVAEB	essential for dipicolinic acid uptake by the developing spore
spoVAF	essential for dipicolinic acid uptake by the developing spore
spoVB	spore cortex synthesis
spoVD	penicillin-binding protein
spoVE	factor for spore cortex peptidoglycan synthesis
spoVS	regulator required for dehydratation of the spore core and assembly of the coat
spoVT	transcriptional positive and negative regulator of σ^G-dependent genes
spsJ	putative dTDP-glucose 4,6-dehydratase/spore coat polysaccharide synthesis
sspB	small acid-soluble spore protein (beta-type SASP)
sspD	small acid-soluble spore protein (alpha/beta-type SASP)
yabG	sporulation-specific protease
yabP	spore protein involved in the shaping of the spore coat
ybaN	polysaccharide deacetylase involved in sporulation
ydfS	hypothetical protein
yfnG	putative sugar-phosphate cytidylyltransferase
ykvU	spore membrane protein involved in germination
ylbJ	putative factor required for spore cortex formation
ymxH	hypothetical protein
yndD	putative spore germination protein
yndF	putative spore germination lipoprotein
yqfD	stage IV sporulation protein
yrbG	hypothetical protein
ytfJ	hypothetical protein
ytlC	putative ABC transporter component, ATP-binding
ytlD	putative permease of ABC transporter
ytvI	putative permease

ancient spore-forming program that was present in a common ancestor.

We wanted to examine the core sporulation gene profiles of the potential endospore-formers in more detail to determine if we could predict endospore formation in these organisms. If their genomes specifically contained the genes necessary for initiation of the spore developmental program, the core regulatory network and engulfment, it is possible these bacteria could carry out the process given the correct conditions. According to our results, we predict *T. potens, A. arabaticum, S. wolfei, S. lipocalidus, T. oceanii, N. thermophiles* and *E. harbinense* could produce endospores while *S. thermophilum, A. degensii, R. albus*, the seven *Caldicellulosiruptor* spp. and the two *Eubacterium* spp. cannot.

The genus *Thermincola* was identified in 2005 (44) and members of this genus have been used in microbial fuel cells (27, 51), but little else has been published about these organisms. *Thermincola potens* has a core sporulation gene profile very similar to that of *Moorella thermoacetica*, a known spore-former (7). It contains homologs of sporulation initiation protein and master regulator Spo0A, the full set of sporulation sigma factors and the genes involved in their regulatory network, and engulfment. It seems likely that either spores have not been seen yet in these recently isolated bacteria or more subtle genetic changes than loss of entire genes prevent formation of mature spores.

Acetohalobium arabaticum and *N. thermophilus* (28, 43) are halophiles and producing a highly resistant spore might be a useful adaptation for surviving in a hypersaline environment. *A. arabaticum* has not been extensively studied but it is an acetate-producing bacterium and the only known member of the genus *Acetohalobium* (43). The core sporulation gene profile of *A. arabaticum* is most similar to *C. difficile*. Although it is missing a homolog of *bofA*, which codes for a protein involved in activating σ^K, many other clostridia also lack a *bofA* homolog and

the pathway leading to active σ^K appears highly divergent in the clostridia. Also, *A. arabaticum* only contains two small-acid soluble proteins (SASPs) out of five examined. These proteins are not crucial to the development of a forespore but do aid in the protection of spore DNA (38, 39). The alkalithermophile *N. thermophilus* has been studied due to its ability to survive in a wide range of environmental extremes (54). Although spores have not been observed in *N. thermophilus*, our results indicate it has a full complement of sporulation genes.

Syntrophomonas wolfei and *Syntrophothermus lipocalidus* are closely related organisms important in syntrophic fatty acid degradation, a thermodynamically unfavorable process (10, 42). For this process to occur, a second species is present, usually a methanogen, that keeps concentrations of degradation products like H_2 and formate at a low level. Thus, these organisms could be used for bioremediation and decreasing greenhouse gas emissions. Our results are in agreement with speculations that *S. wolfei* is an endospore former due to the high number of sporulation genes in its genome (42). The only significant gene that appears to be missing is *spoIIIE*, which codes for an FtsK homolog (41) important for DNA translocation into the spore (52) and membrane fusion during engulfment (26, 40). Clostridia may have other mechanisms of translocating DNA as other clostridia, such as *Clostridium phytofermentans*, *Clostridium saccharolyticum* and *C. difficile*, lack obvious homologs of *spoIIIE*. Although *S. lipocalidus* was tested for endospore formation under a variety of conditions, including both high and low temperatures, spores were never observed (37). The most similar sporulation gene profile to both *S. wolfei* and *S. lipocalidus* from a known spore-former is *Clostridium botulinum*.

Thermosediminibacter oceani was isolated from a core sample of deep sea sediments in 2005 and has the ability to ferment a variety of polysaccharides but has not been well studied (25). The lower limit for growth of *T. oceani* is 50°C. According to our results, *T. oceani*

contains all the genes necessary to form an endospore. Although spores have not been observed in cultures of *T. oceani* (25), the production of a spore would be useful for dispersal of *T. oceani* and to survive lower temperatures.

Identified in 2006, *E. harbinense* has been studied for its ability to produce hydrogen and ethanol during sugar fermentation (16, 53). It has a sporulation gene profile most similar to those of *Clostridium* spp., and appears to have almost all the genes necessary to produce a spore. However, the Spo0A homolog of *E. harbinense* appears the most divergent from the *B. subtilis* Spo0A as it has only a 41% amino acid identity compared to 50% or more for homologs from known endospore-formers. Also, the *E. harbinense* genome lacks a clear homolog of *spoIIR*, the protein responsible for signaling from the forespore to the mother cell for the activation of σ^E (19). A few endospore-formers like *Alkaliphilus oremlandii* and *Halothermothrix orenii* lack homologs of *spoIIR*, but homologs are found in almost all others. While it is possible that these differences prevent endospore formation in *E. harbinense*, the high conservation of other key sporulation genes leads us to speculate that it may have the genetic capacity necessary to produce a spore.

All of the other potential endospore-formers lack a combination of homologs of key genes needed for endospore formation. In addition, their homologs of the sporulation-specific sigma factors and/or *spo0A* are less similar to the *B. subtilis* genes than those of known endospore-formers. The *Caldicellulosiruptor* spp., *Eubacterium* spp., and *R. albus* are also missing two key components for engulfment, *spoIIM* and *spoIIP* as well as *spoIIIAB*, which is conserved in all but one known endospore-former. Both *S. thermophilum* and *R. albus* are missing *spoIIGA*, which is important for σ^E activation. A *spoIIR* homolog and the genes required for σ^K activation are missing from *A. degensii*. Missing a single genetic component like

spoIIR might not abolish sporulation, but we believe multiple defects presents a more significant barrier that would likely stop the cell from producing a spore.

Conclusions. The results presented here view the evolution of endospore formation from the perspective of *B. subtilis*. Although hundreds of genes are involved in the process of forming a dormant spore only a subset appear to be conserved in all endospore-formers. These genes likely represent a core sporulation program preserved from a common ancestral endospore-former. Using these genes, can we predict spore formation in other organisms as we have attempted to do here? While the absence of some genes might be a clear indication of a non-sporulating phenotype, others are not as obvious. How many genes need to be absent before endospore formation is abolished? We already know that endospore formation can be used for more than survival as others like *M. polyspora* use it for reproduction (5). What might be the next step in the evolution of the process? *Epulopiscium* sp. type B, a fish symbiont, likely uses a modified sporulation program to produce multiple intracellular offspring that are not dormant like endospores (3, 4, 8, 29). Binary fission has never been observed in this organism, making intracellular offspring formation its only strategy for reproduction. While this research provides a good foundation for understanding the conservation of endospore formation in the Firmicutes, it still provides a very *Bacillus*-centric view. Sporulation programs from other Firmicutes must be examined in detail to provide a more unbiased view of the evolution of this process.

REFERENCES

1. **Adams, D. G., and P. S. Duggan.** 1999. Tansley Review No. 107 Heterocyst and Akiniete Differentiation in Cyanobacteria. New Phytologist. **144:**3-33.
2. **Angert, E. R.** 2005. Alternatives to binary fission in bacteria. Nat. Rev. Microbiol. **3:**214-24.
3. **Angert, E. R., A. E. Brooks, and N. R. Pace.** 1996. Phylogenetic analysis of *Metabacterium polyspora*: clues to the evolutionary origin of daughter cell production in *Epulopiscium* species, the largest bacteria. J. Bacteriol. **178:**1451-6.
4. **Angert, E. R., and K. D. Clements.** 2004. Initiation of intracellular offspring in *Epulopiscium*. Mol. Microbiol. **51:**827-35.
5. **Angert, E. R., and R. M. Losick.** 1998. Propagation by sporulation in the guinea pig symbiont *Metabacterium polyspora*. Proc. Natl. Acad. Sci. U. S. A. **95:**10218-23.
6. **Ben-Yehuda, S., and R. Losick.** 2002. Asymmetric cell division in *B. subtilis* involves a spiral-like intermediate of the cytokinetic protein FtsZ. Cell **109:**257-66.
7. **Byrer, D. E., F. A. Rainey, and J. Wiegel.** 2000. Novel strains of *Moorella thermoacetica* form unusually heat-resistant spores. Arch. Microbiol. **174:**334-9.
8. **Clements, K. D., D. C. Sutton, and J. H. Choat.** 1989. Occurrence and characteristics of unusual protistan symbionts from surgeonfishes (Acanthuridae) of the Great Barrier Reef, Australia. Marine Biol. **102:**403-12.
9. **de Hoon, M. J., P. Eichenberger, and D. Vitkup.** 2010. Hierarchical evolution of the bacterial sporulation network. Curr. Biol. **20:**R735-45.

10. **Djao, O. D., X. Zhang, S. Lucas, A. Lapidus, T. G. Del Rio, M. Nolan, H. Tice, J. F. Cheng, C. Han, R. Tapia, L. Goodwin, S. Pitluck, K. Liolios, N. Ivanova, K. Mavromatis, N. Mikhailova, G. Ovchinnikova, A. Pati, E. Brambilla, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, M. Rohde, J. Sikorski, S. Spring, M. Goker, J. C. Detter, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, and H. P. Klenk.** 2010. Complete genome sequence of *Syntrophothermus lipocalidus* type strain (TGB-C1). Stand. Genomic. Sci. **3:**268-75.

11. **Eichenberger, P., M. Fujita, S. T. Jensen, E. M. Conlon, D. Z. Rudner, S. T. Wang, C. Ferguson, K. Haga, T. Sato, J. S. Liu, and R. Losick.** 2004. The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. PLoS Biol. **2:**e328.

12. **Eichenberger, P., S. T. Jensen, E. M. Conlon, C. van Ooij, J. Silvaggi, J. E. Gonzalez-Pastor, M. Fujita, S. Ben-Yehuda, P. Stragier, J. S. Liu, and R. Losick.** 2003. The σ^E regulon and the identification of additional sporulation genes in *Bacillus subtilis*. J. Mol. Biol. **327:**945-72.

13. **Errington, J.** 2003. Regulation of endospore formation in *Bacillus subtilis*. Nat. Rev. Microbiol. **1:**117-26.

14. **Felsenstein, J.** 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics **5:**164-6.

15. **Fonknechten, N., S. Chaussonnerie, S. Tricot, A. Lajus, J. R. Andreesen, N. Perchat, E. Pelletier, M. Gouyvenoux, V. Barbe, M. Salanoubat, D. Le Paslier, J. Weissenbach, G. N. Cohen, and A. Kreimeyer.** 2010. *Clostridium sticklandii*, a specialist in amino acid degradation:revisiting its metabolism through its genome sequence. BMC Genomics **11:**555.

16. **Guo, W. Q., N. Q. Ren, X. J. Wang, W. S. Xiang, J. Ding, Y. You, and B. F. Liu.** 2009. Optimization of culture conditions for hydrogen production by *Ethanoligenens harbinense* B49 using response surface methodology. Bioresour. Technol. **100:**1192-6.

17. **Hilbert, D. W., and P. J. Piggot.** 2004. Compartmentalization of gene expression during *Bacillus subtilis* spore formation. Microbiol. Mol. Biol. Rev. **68:**234-62.

18. **Hoch, J. A.** 1998. Initiation of bacterial development. Curr. Opin. Microbiol. **1:**170-4.

19. **Hofmeister, A. E., A. Londono-Vallejo, E. Harry, P. Stragier, and R. Losick.** 1995. Extracellular signal protein triggering the proteolytic activation of a developmental transcription factor in *B. subtilis*. Cell **83:**219-26.

20. **Kay, D., and S. C. Warren.** 1968. Sporulation in *Bacillus subtilis*. Morphological changes. Biochem. J. **109:**819-24.

21. **Kroos, L.** 2007. The *Bacillus* and *Myxococcus* developmental networks and their transcriptional regulators. Annu. Rev. Genet. **41:**13-39.

22. **Labbé, R. G.** 2005. Sporulation (Morphology) of Clostridia, p. 647-658. *In* P. Dürre (ed.), Handbook on Clostridia. CRC Press, Boca Raton, FL.

23. **Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins.** 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23:**2947-8.

24. **Lechat, P., L. Hummel, S. Rousseau, and I. Moszer.** 2008. GenoList: an integrated environment for comparative analysis of microbial genomes. Nucleic Acids Res. **36:**D469-74.

25. **Lee, Y. J., I. D. Wagner, M. E. Brice, V. V. Kevbrin, G. L. Mills, C. S. Romanek, and J. Wiegel.** 2005. *Thermosediminibacter oceani* gen. nov., sp. nov. and *Thermosediminibacter litoriperuensis* sp. nov., new anaerobic thermophilic bacteria isolated from Peru Margin. Extremophiles **9:**375-83.

26. **Liu, N. J., R. J. Dutton, and K. Pogliano.** 2006. Evidence that the SpoIIIE DNA translocase participates in membrane fusion during cytokinesis and engulfment. Mol. Microbiol. **59:**1097-113.

27. **Mathis, B. J., C. W. Marshall, C. E. Milliken, R. S. Makkar, S. E. Creager, and H. D. May.** 2008. Electricity generation by thermophilic microorganisms from marine sediment. Appl. Microbiol. Biotechnol. **78:**147-55.

28. **Mesbah, N. M., D. B. Hedrick, A. D. Peacock, M. Rohde, and J. Wiegel.** 2007. *Natranaerobius thermophilus* gen. nov., sp. nov., a halophilic, alkalithermophilic bacterium from soda lakes of the Wadi An Natrun, Egypt, and proposal of Natranaerobiaceae fam. nov. and Natranaerobiales ord. nov. Int. J. Syst. Evol. Microbiol. **57:**2507-12.

29. **Miller, D. A., J. H. Choat, K. D. Clements, and E. R. Angert.** 2011. The *spoIIE* homolog of *Epulopiscium* sp. type B is expressed early in intracellular offspring development. J. Bacteriol. **193:**2642-6.

30. **Molle, V., M. Fujita, S. T. Jensen, P. Eichenberger, J. E. Gonzalez-Pastor, J. S. Liu, and R. Losick.** 2003. The Spo0A regulon of *Bacillus subtilis*. Mol. Microbiol. **50:**1683-701.

31. **Ohnishi, Y., J. W. Seo, and S. Horinouchi.** 2002. Deprogrammed sporulation in *Streptomyces*. FEMS Microbiol. Lett. **216:**1-7.

32. **Onyenwoke, R. U., J. A. Brill, K. Farahi, and J. Wiegel.** 2004. Sporulation genes in members of the low G+C Gram-type-positive phylogenetic branch (Firmicutes). Arch. Microbiol. **182:**182-92.

33. **Paredes, C. J., K. V. Alsaker, and E. T. Papoutsakis.** 2005. A comparative genomic view of clostridial sporulation and physiology. Nat. Rev. Microbiol. **3:**969-78.

34. **Ravagnani, A., K. C. Jennert, E. Steiner, R. Grunberg, J. R. Jefferies, S. R. Wilkinson, D. I. Young, E. C. Tidswell, D. P. Brown, P. Youngman, J. G. Morris, and M. Young.** 2000. Spo0A directly controls the switch from acid to solvent production in solvent-forming clostridia. Mol. Microbiol. **37:**1172-85.

35. **Rothfield, L., S. Justice, and J. Garcia-Lara.** 1999. Bacterial cell division. Annu. Rev. Genet. **33:**423-48.

36. **Ryter, A.** 1965. Etude morphologique de la sporulation de *Bacillus subtilis*. Ann. Inst. Pasteur. **108:**40-60.

37. **Sekiguchi, Y., Y. Kamagata, K. Nakamura, A. Ohashi, and H. Harada.** 2000. *Syntrophothermus lipocalidus* gen. nov., sp. nov., a novel thermophilic, syntrophic, fatty-acid-oxidizing anaerobe which utilizes isobutyrate. Int. J. Syst. Evol. Microbiol. **50 Pt 2:**771-9.

38. **Setlow, B., K. A. McGinnis, K. Ragkousi, and P. Setlow.** 2000. Effects of major spore-specific DNA binding proteins on *Bacillus subtilis* sporulation and spore properties. J. Bacteriol. **182:**6906-12.

39. **Setlow, P.** 2007. I will survive: DNA protection in bacterial spores. Trends Microbiol. **15:**172-80.

40. **Sharp, M. D., and K. Pogliano.** 2003. The membrane domain of SpoIIIE is required for membrane fusion during *Bacillus subtilis* sporulation. J. Bacteriol. **185:**2005-8.

41. **Sharpe, M. E., and J. Errington.** 1995. Postseptational chromosome partitioning in bacteria. Proc. Natl. Acad. Sci. U. S. A. **92:**8630-4.

42. **Sieber, J. R., D. R. Sims, C. Han, E. Kim, A. Lykidis, A. L. Lapidus, E. McDonnald, L. Rohlin, D. E. Culley, R. Gunsalus, and M. J. McInerney.** 2010. The genome of *Syntrophomonas wolfei*: new insights into syntrophic metabolism and biohydrogen production. Environ. Microbiol.

43. **Sikorski, J., A. Lapidus, O. Chertkov, S. Lucas, A. Copeland, T. Glavina Del Rio, M. Nolan, H. Tice, J. F. Cheng, C. Han, E. Brambilla, S. Pitluck, K. Liolios, N. Ivanova, K. Mavromatis, N. Mikhailova, A. Pati, D. Bruce, C. Detter, R. Tapia, L. Goodwin, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y. J. Chang, C. D. Jeffries, M. Rohde, M. Goker, S. Spring, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, and H. P. Klenk.** 2010. Complete genome sequence of *Acetohalobium arabaticum* type strain (Z-7288). Stand. Genomic. Sci. **3:**57-65.

44. **Sokolova, T. G., N. A. Kostrikina, N. A. Chernyh, T. V. Kolganova, T. P. Tourova, and E. A. Bonch-Osmolovskaya.** 2005. *Thermincola carboxydiphila* gen. nov., sp. nov., a novel anaerobic, carboxydotrophic, hydrogenogenic bacterium from a hot spring of the Lake Baikal area. Int. J. Syst. Evol. Microbiol. **55:**2069-73.

45. **Stadtman, T. C., and L. S. McClung.** 1957. *Clostridium sticklandii* nov. spec. J. Bacteriol. **73:**218-9.

46. **Steil, L., M. Serrano, A. O. Henriques, and U. Volker.** 2005. Genome-wide analysis of temporally regulated and compartment-specific gene expression in sporulating cells of *Bacillus subtilis*. Microbiology **151:**399-420.

47. **Stephenson, K., and J. A. Hoch.** 2002. Evolution of signalling in the sporulation phosphorelay. Mol. Microbiol. **46:**297-304.

48. **Stragier, P.** 2002. A gene odyssey: exploring the genomes of endospore-forming bacteria, p. 519-525. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and its closest relatives. ASM Press, Washington D.C.

49. **Stragier, P., and R. Losick.** 1996. Molecular genetics of sporulation in *Bacillus subtilis*. Annu. Rev. Genet. **30:**297-41.

50. **Wang, S. T., B. Setlow, E. M. Conlon, J. L. Lyon, D. Imamura, T. Sato, P. Setlow, R. Losick, and P. Eichenberger.** 2006. The forespore line of gene expression in *Bacillus subtilis*. J. Mol. Biol. **358:**16-37.

51. **Wrighton, K. C., P. Agbo, F. Warnecke, K. A. Weber, E. L. Brodie, T. Z. DeSantis, P. Hugenholtz, G. L. Andersen, and J. D. Coates.** 2008. A novel ecological role of the Firmicutes identified in thermophilic microbial fuel cells. ISME J. **2:**1146-56.

52. **Wu, L. J., and J. Errington.** 1994. *Bacillus subtilis* SpoIIIE protein required for DNA segregation during asymmetric cell division. Science **264:**572-5.

53. **Xing, D., N. Ren, Q. Li, M. Lin, A. Wang, and L. Zhao.** 2006. *Ethanoligenens harbinense* gen. nov., sp. nov., isolated from molasses wastewater. Int. J. Syst. Evol. Microbiol. **56:**755-60.

54. **Zhao, B., N. M. Mesbah, E. Dalin, L. Goodwin, M. Nolan, S. Pitluck, O. Chertkov, T. S. Brettin, J. Han, F. W. Larimer, M. L. Land, L. Hauser, N. Kyrpides, and J. Wiegel.** 2011. Complete Genome Sequence of the Anaerobic, Halophilic Alkalithermophile *Natranaerobius thermophilus* JW/NM-WN-LF. J. Bacteriol. **193:**4023-4.

CHAPTER 2

THE GENOMIC BASIS FOR INTRACELLULAR OFFSPRING FORMATION IN THE BACTERIUM *EPULOPISCIUM*

ABSTRACT

Epulopiscium sp. type B, a large intestinal symbiont of the surgeonfish *Naso tonganus*, does not reproduce by binary fission. Instead, it forms multiple intracellular offspring. Offspring production in *Epulopiscium* shares morphological features with the survival strategy of endospore formation in the Firmicutes. We hypothesize that the two processes are related and share molecular mechanisms that are responsible for the observed morphological similarities. To test this, we sequenced to draft quality, the genome of *Epulopiscium* sp. type B and used it, along with the complete genome of its close, endospore-forming relative *Cellulosilyticum lentocellum*, to identify homologs of sporulation genes characterized in *Bacillus subtilis*. Of the 147 highly conserved *Bacillus subtilis* sporulation genes used in this analysis, we found 57 homologs in the *Epulopiscium* genome and 87 homologs in the *C. lentocellum* genome. Genes coding for components of the central regulatory network which govern the expression of forespore and mother-cell-specific sporulation genes and the machinery used for engulfment appear best conserved, indicating that these may be ancestral sporulation mechanisms. Low conservation of genes expressed late in endospore formation, particularly those that confer resistance properties and germinant receptors, suggest that *Epulopiscium* has lost the ability to form a mature spore. Our findings provide a basis for understanding the evolutionary relationship between intracellular offspring production and endospore formation.

INTRODUCTION

Endospore formation is an ancient and complex developmental process exhibited only by certain bacteria within the Firmicutes (26, 68). Endospores endure environmental assaults that would kill most other bacterial cells. They can survive prolonged periods of insufficient nutrients, moderate levels of organic solvents, exposure to phage, extremes in pH, proteases and cell wall degrading enzymes, freezing, desiccation and excessive heat or radiation (67, 80). This form of sporulation preserves the genome in a remarkably dispersible and dormant cell type that can resume vegetative growth when the environment improves. While most sporulating Firmicutes produce a single endospore, some have the ability to produce multiple endospores (4). For example, *Clostridium oceanicum* regularly forms two endospores, one at each end of the mother cell (84). Others include the Segmented Filamentous Bacteria, a group of uncultivated inhabitants of the intestinal tract of animals (49). These multicellular filaments live attached to the lining of the small intestine, and to disperse or reposition itself in the gut, each cell in a filament forms an endospore containing two cells or alternatively produces two non-dormant intracellular offspring (19, 49).

Other lineages within the Firmicutes use multiple endospore formation as a reproductive strategy. The guinea pig intestinal symbiont *Metabacterium polyspora* may undergo binary fission but the regular formation of multiple endospores, up to nine from a single mother cell, is a significant form of reproduction (7, 75). The life history of *M. polyspora* may be selecting for this unusual mode of reproduction, which could improve survival as the bacteria cycle in and out of the host gastrointestinal tract (7). A large and diverse group of surgeonfish intestinal symbionts related to *M. polyspora* display an array for reproductive modes that employ binary fission and/or sporulation (22). The type C *Epulopiscium*-like fish intestinal symbionts, for

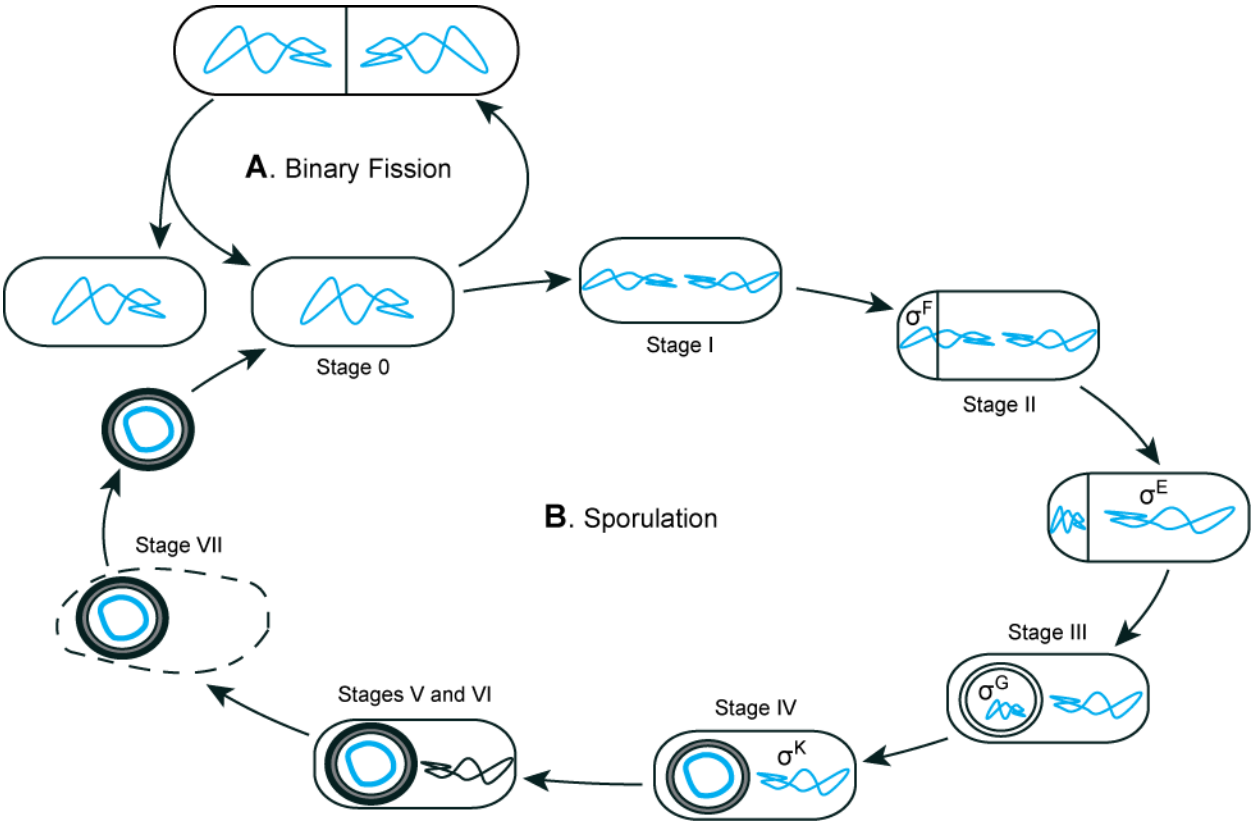


Figure 2.1. The *B. subtilis* life cycle and stages of sporulation. A) In a favorable environment, *B. subtilis* undergoes growth and division. B) When nutrient limitations become critical, the cell may develop an endospore. Shown here are the morphological stages described for sporulation. The temporal and spatial activation of the four sporulation-specific sigma factors is shown. DNA is shown in blue. The grey circle around the forespore indicates the cortex. The thick black circle around the forespore at stage V and beyond represents the spore coat.

example, rely solely on the formation of two endospores for reproduction and appear to have abandoned binary fission (36). The largest members of this group of symbionts, *Epulopiscium* spp. type A and type B, seem to have taken this process one step further and reproduce by the daily production of two or more intracellular offspring that are not dormant (5, 22). The close phylogenetic relationship between multiple endospore formers and lineages that produce non-dormant intracellular offspring, and the morphological changes shared between these processes, suggest that the latter developmental process is related to endospore formation (5).

The stages of endospore formation (Figure 2.1) are described in the model organism *Bacillus subtilis* (48, 76) and these morphological transitions appear conserved in other endospore formers (35). Cells that exhibit no overt signs of sporulation are defined as stage 0. After initiation of sporulation, the chromosome replicates and replication origins become tethered to opposite poles of the cell. This unusual nucleoid conformation is called the axial filament and these cells are said to be in stage I (15). Instead of dividing at the midcell, the sporulating cell divides near one pole, producing the forespore and larger mother cell, which marks stage II. Division traps approximately one-third of the origin proximal region of one of the chromosomes in the forespore (96). The rest of the chromosome, still within the mother cell, is translocated into the forespore so that the spore contains a complete genome (12). Enzymatic degradation of peptidoglycan between the mother cell and forespore results in curvature of the septum (10, 20). The mother-cell membrane then wraps around the forespore to completely engulf the forespore, which marks stage III (13). In stage IV, a modified peptidoglycan called the cortex is synthesized in the space between the mother-cell and forespore membranes. In stage V, a complex proteinaceous coat is applied to the developing spore (32, 42). Stage VI is defined as endospore maturation, when the spore gains many of its resistance traits (67). Lastly,

the mother cell lyses to release the mature spore, in stage VII.

The sequential activation of stage-specific transcription factors controls the proper timing and location of gene expression to ensure the progression of cell-specific developmental events. In *B. subtilis*, a network of kinases and phosphatases conveys information about intracellular and extracellular conditions to the phosphorelay, which ultimately determines the phosphorylation state of the transcription factor Spo0A (14, 38, 93). While Spo0A is considered the master regulator of sporulation, it also regulates a number of alternative cellular reactions to environmental change (55). In its active form, Spo0A~P either directly or indirectly affects the transcription of more than 500 genes (65). Spo0A activation is essential for entry into sporulation.

After asymmetric division, gene expression is regulated by four sporulation-specific sigma factors: σ^F, σ^E, σ^G, and σ^K (Figure 1) (43, 56). σ^F and σ^G are activated only in the forespore while σ^E and σ^K are activated only in the mother cell. Both σ^F and σ^E are expressed prior to asymmetric division (39), but σ^F is held inactive in a complex with two peptides of its anti-sigma factor SpoIIAB (29) and σ^E is synthesized as an inactive pro-peptide (51). The forespore sigma, σ^F, is the first to be activated. SpoIIE phosphorylates SpoIIAA (the σ^F anti-anti-sigma factor), which binds SpoIIAB leading to the release of σ^F (8, 21, 28). SpoIIR, part of the σ^F regulon (47), is produced in the forespore and inserted into the sporulation septum, where it activates SpoIIGA in the mother cell (44). SpoIIGA then cleaves pro-σ^E thus releasing mature σ^E to the cytoplasm (45). Likewise, the late-sporulation sigma factors, σ^G and σ^K, are not immediately functional when expressed and activation of each entails factor-specific intracellular signaling cascades and release mechanisms (16, 17, 24, 34, 50, 61, 70, 74).

Tighter control of particular genes in each regulon is provided by additional transcription factors (11, 30, 41, 46, 97) that form both coherent and incoherent feed-forward loops with their associated sigma factor (26). Coherent feed-forward loops occur when a sigma factor regulates expression of a gene and then combines with that gene product to up-regulate more genes. Incoherent feed-forward loops come about when such a combination leads to the down-regulation of additional genes (26, 30). For example, *rsfA* and *spoIIR* are both expressed from σ^F promoters (46, 47). However, RsfA combined with σ^F turns off transcription of *spoIIR*, so only a brief burst of *spoIIR* expression is seen immediately following asymmetric division. Such feed-forward loops allow the cell to modulate the timing, duration and location of expression of subsets of genes within a regulon. The combination of these central transcriptional regulatory mechanisms ensures the appropriate expression of more than 700 genes that help create a mature endospore.

The *B. subtilis* model can serve as a foundation for exploring mechanistic modifications required to support the formation of multiple endospores or intracellular offspring (5). For example, *Epulopiscium* sp. type B are intestinal symbionts of the unicornfish *Naso tonganus* and can reach lengths of 200-300 µm and widths of 50-60 µm (6, 22). While the production of endospores has been observed in related morphotypes (36), *Epulopiscium* sp. type B does not produce endospores and does not reproduce by binary fission (Figure 2.2). Instead, each cell forms two or more intracellular offspring in a process that repeats daily (6, 63, 95). The formation of these offspring has been described in stages that parallel stages of endospore formation. Stage 0 cells have large offspring (daughter cells) that show no signs of the initiation of the next generation of offspring (i.e. granddaughter cells). Stage I is defined as offspring cells that have coalesced DNA at the poles. Stage II cells have straight polar septa, but not all polar

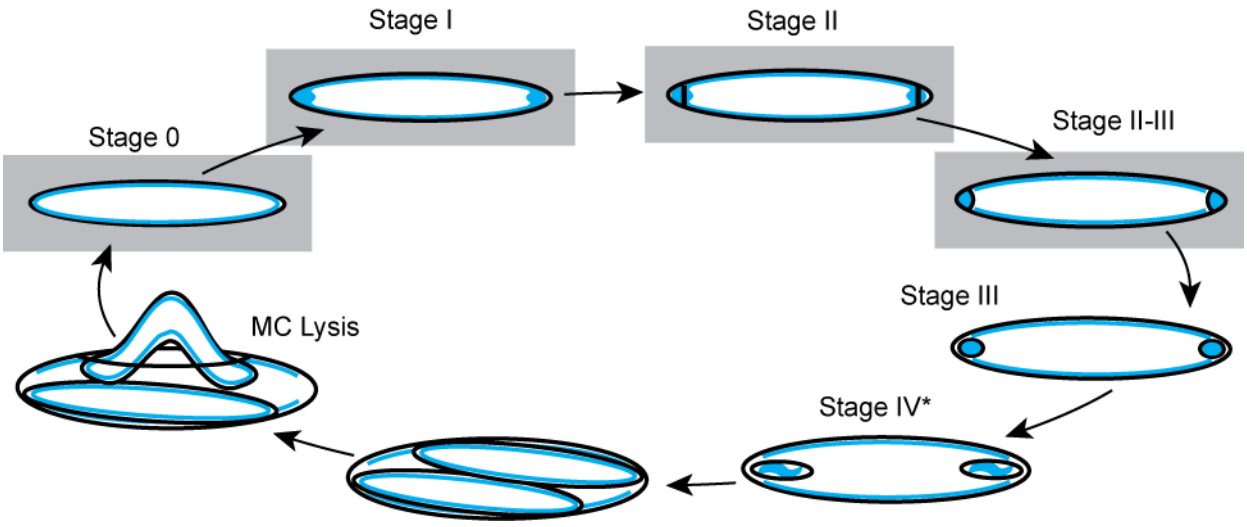


Figure 2.2. The *Epulopiscium* sp. type B life cycle. Earliest stages of offspring development are based on the similar morphological transitions described for sporulation in *B. subtilis*. See text for a detailed explanation of the process. Offspring frequently initiate the next round of reproduction prior to exiting the mother cell. Those stages that are seen in offspring still within their mother cell are highlighted with grey boxes. DNA is shown in blue.

DNA is inside the newly formed offspring. Stage II-III includes cells with curved polar septa, indicating the start of polar cell engulfment, and all polar DNA has been translocated into the offspring. Stage III cells contain small, fully engulfed offspring, with a length-to-width ratio of less than 2:1 while the offspring in stage IV* cells have a ratio greater than 2:1. After engulfment, the two processes diverge and changes occurring in *Epulopiscium* are not well understood so later stages are not indicated in this model. Offspring continue to grow until they fill the mother-cell cytoplasm. In time, a tear forms in the mother-cell envelope and the offspring are released. Previous work has shown that the division protein FtsZ localizes to the poles of an *Epulopiscium* cell in a similar way to FtsZ in *B. subtilis* during endospore formation (6). Putative homologs for SpoIIE, SpoIIAA, SpoIIAB and σ^F also have been identified in the

Epulopiscium genome (63). Moreover, the expression pattern of *spoIIE* in *Epulopiscium* during offspring formation (63) is very similar to that seen in sporulating *B. subtilis* (40). These results suggest that *Epulopiscium* uses cell-specific activation of alternative sigma factors in offspring development. Here we sequenced and examined the draft genome of *Epulopiscium* sp. type B to determine the extent of conservation of the sporulation genetic program. A list of conserved "core" sporulation genes was assembled. BLAST was then used to determine which of the core sporulation genes are conserved in the *Epulopiscium* genome and which are conserved in its spore-forming relative *Cellulosilyticum lentocellum* DSM 5427. As predicted, we found a number of homologs to genes with sporulation-specific functions in *Epulopiscium* and even more of the core genes conserved in *C. lentocellum*. These results begin to define the genetic mechanisms that may be used for offspring development in *Epulopiscium*.

MATERIALS AND METHODS

***Epulopiscium* sample collection and DNA extraction.** *Naso tonganus* were collected by spear on outer reefs in the vicinity of Lizard Island, Great Barrier Reef, Australia. Sections of the gut were removed and the contents were fixed in 80% ethanol. Samples were stored at -20°C upon arrival at the laboratory.

Epulopiscium cells were manually selected from fixed intestinal contents, using a standard Gilson pipettor and a Nikon SMZ-U dissecting microscope. Cell lysis and DNA extraction were performed as previously described (62). Briefly, *Epulopiscium* type B cells were incubated with 100 µg/ml Proteinase K for one hour at 50°C. DNA was extracted with phenol:chloroform and precipitated. After rinsing with 70% ethanol, the DNA was resuspended in TE (10mM Tris, 1 mM EDTA, pH 8) buffer.

Genome sequencing and analysis. A draft genome sequence was generated using a random shotgun approach and paired-end Sanger sequencing. Sequence reads, providing approximately 8-fold coverage of the estimated 4 Mb genome were assembled using a combination of the Celera Assembler (66) and TIGR assembler (90). To improve assembly, reads were first sorted based on G+C content and only reads that had 23 – 53% G+C were retained. The resulting reads were assembled again and open reading frames (ORFs) predicted using GLIMMER (27). A BLAST (1) analysis of predicted ORFs was used to remove non-bacterial sequences. This strategy yielded 92 large contigs 12 kb to 119 kb in length. These contigs are deposited in GenBank under the accession number NZ_ABEQ01000000.

Draft genome assessment of the 92 large contigs was performed using two approaches. First, the total number of rRNA and tRNA genes were predicted using RNAmmer (52) and tRNAscan-SE (57), respectively. Second, a clusters of orthologous analysis (COG) (91) was

performed using the *Epulopiscium* predicted proteome. Each protein was compared against a local COG database obtained from NCBI (ftp://ftp.ncbi.nih.gov/pub/mmdb/cdd/, accessed: 06/23/2011) using RPSBLAST (58) and the total number of proteins in each COG category was tabulated and represented as a percent of the number of COG-annotated proteins in the genome. This analysis was also performed for all sequenced genomes belonging to the phylum Firmicutes and genus *Clostridium* in the complete sequenced microbial genome collection (http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi, accessed: 06/23/2011).

Generation of a core sporulation gene list. An initial list of sporulation genes was created based on four studies describing the regulons of the four sporulation-specific sigma factors and the transcription factor Spo0A from *B. subtilis* (30, 65, 85, 94). Other genes in GenoList with putative sporulation functions in *B. subtilis* 168 were added to this list (53). Any gene that codes for a protein with a known vegetative growth function was removed. The list was refined further by BLASTp (http://blast.ncbi.nlm.nih.gov/Blast.cgi) searches to determine the distribution of similar proteins in the GenBank non-redundant protein database. Multiple GenBank searches were performed and limited to phylum Firmicutes, class Clostridia, and if necessary *Clostridium* spp. and *Bacillus* spp. Proteins represented in 15 strains of endospore-forming bacteria in both class Bacilli and class Clostridia with greater than or equal to 70% query coverage and 20% identity were retained on the list. Proteins with top hits for putative homologs in two or more species within the non-spore-forming genera *Listeria*, *Streptococcus*, *Lactobacillus*, *Staphylococcus*, *Enterococcus*, and *Peptostreptococcus*, indicating a vegetative function, were removed. Several key proteins that were first identified as essential to sporulation (Spo0A, Spo0J, Soj, SpoIIIE and SpoIIIJ) were added back to the list.

Search for core proteins coded for in the *Epulopiscium* and *Cellulosilyticum lentocellum*

genomes. All contigs and well-represented single reads from the *Epulopiscium* sp. type B draft genome were linked together to form a pseudomolecule. Coordinates for all junctions between contigs and single reads were entered into an Excel file for easy reference. The pseudomolecule and the complete *C. lentocellum* DSM 5427 genome (64) were searched for homologs of core proteins using tBLASTn and BLASTp, respectively. For each query, the top hits in the *Epulopiscium* or *C. lentocellum* genome were compared to *Bacillus* and *Clostridium* sequences in the non-redundant database at NCBI using BLASTp. If reverse-BLAST searches recovered a top hit that was different from the *B. subtilis* core protein used in the initial BLAST analysis, the protein was eliminated from the list. By this method, additional putative homologs with weak hits (below the original cutoff values) were identified. Operon structure was an additional criterion used to identify homologs. Core sporulation protein homologs identified in *C. lentocellum* were then used to search the *Epulopiscium* genome using tBLASTn.

Confirmation of *dacF, gpr-spoIIP, spoVT, sspC/F, ybdM and yyaC*. Primers were developed to amplify sporulation genes identified in *Epulopiscium* genome that were located in small contigs or singletons or to link genes in operons predicted from synteny in the genomes of other spore formers (Table 2.1). PCR was performed using standard reaction conditions, and products were cloned into the pCR 2.1 TOPO vector (Invitrogen) following the protocol provided by the manufacturer. Sequences of the clones were determined using Big Dye Terminator chemistry and an Applied Biosystems Automated 3730 DNA analyzer, performed at the Cornell Life Sciences Core Laboratories Center and analyzed with Geneious Pro 5.4.2 (Biomatters). The sequences for the *gpr-spoIIP* operon (accession number JN402987), *dacF* (JN402985), *spoVT* (JN402986), *sspC/F* (JN402984), *ybdM* (JN402982), and *yyaC* (JN402983) are available from GenBank.

Table 2.1. Primers used in this study

DESIGNATION	SEQUENCE (5′ TO 3′)
GprendF	ATAGACGCATTAGGAGCACG
SpoIIPbegR	GCTTAGCGGACTTTGTATCACC
EpuloGprF	GAGAACATTGGTATTACAGGCG
EpuloGprR	GCTTGCATATATCACCTCCTTG
EpuloSpoIIPF	CATTGCTGTTCACCCAGGTA
EpuloSpoIIP859R	GCAGTAACCTTAGACGCA
EpuloSpoIIP684F	CAAAGTATGGGCTAAATGTATTGC
EpuloSpoIIPR	CAAAACAACAGACATCACCG
EpuloDacFF	GAGCCCCTGATTGTAACATTT
EpuloDacFR1	GGTTAATCCAAATTCACTTTCGCC
EpuloSpoVTF	TTTAGTATTATCAAGAGAAAAACAGCAT
EpuloSpoVTR	TGAACATTTGTCAAGATATAAATGCAA
EpuloSspC/FF	CTCCAAAATAATTTAGGAATATTGTCC
EpuloSspC/FR	TACACAGAAGTACCCCTTTGC
EpuloYbdMF	GAGTGGTTCTTTTAGTGGCTCTG
EpuloYbdMR	AACATGACCTCACACTGGCA
EpuloYyaCF	GCGGTGTTTGTTGTAAGTGC
EpuloYyaCR	TACACCCGAAGAATTAAGCA

RESULTS AND DISCUSSION

***Epulopiscium* sp. type B draft genome.** The draft genome of *Epulopiscium* sp. type B was used to assess the conservation of sporulation gene homologs in this bacterium. Although the genome was not assembled completely, the project produced approximately 2.7 Mb of sequence in 92 large contigs ranging from 12 kb to 119 kb in length. The remaining assembled data, of small contigs less than 12 kb and well represented single reads (singletons), comprise another 13.7 Mb. This data set likely represents the *Epulopiscium* genome and may contain some highly repetitive DNA from the fish host or abundant inhabitants of the *N. tonganus* intestinal tract that may have been carried along during *Epulopiscium* cell isolation.

We analyzed the 92 large contigs of the draft genome of *Epulopiscium* using two approaches. First, we identified the total number of rRNAs and tRNAs encoded by the genome. We found four 5S, four 23S, and three 16S rRNA genes in addition to 31 tRNAs covering all amino acids except histidine, phenylalanine, and serine. These values are about one-third less the rRNAs and tRNAs encoded by *Cellulolilyticum lentocellum* DSM 5427, the closest relative with a complete genome sequence (64). We also analyzed the singletons generated as part of the *Epulopiscium* genome-sequencing project and found additional functional RNAs. This resulted in a total of nine 5S, five 16S and five 23S rRNAs, as well as 66 tRNAs covering all amino acids for the *Epulopiscium* draft genome. These numbers should be viewed with caution as some copies may represent overlapping contigs that did not assemble. Second, we performed a COG analysis using the predicted proteome of the 92 large contigs from *Epulopiscium* and compared the percent proteins encoded in each category against a similar analysis of all complete sequenced genomes in the phylum Firmicutes and the genus *Clostridium* (Table 2.2). We found that the majority of categories had a slightly higher percentage of proteins devoted to that

Table 2.2. COG analysis

COG Category	Code	*Epulopiscium* sp. type *B*	*Clostridium*	Firmicutes
Information storage and processing				
Translation, ribosomal structure and biogenesis	**J**	7.57%	5.91%	6.95%
RNA processing and modification	**A**	0.00%	0.00%	0.00%
Transcription	**K**	5.35%	9.19%	8.05%
Replication, recombination and repair	**L**	4.17%	5.43%	6.20%
Chromatin structure and dynamics	**B**	0.00%	0.03%	0.03%
Cellular processes and signaling				
Cell cycle control, cell division, chromosome partitioning	**D**	1.94%	1.24%	1.29%
Nuclear structure	**Y**	0.00%	0.00%	0.00%
Defense mechanisms	**V**	1.60%	2.81%	2.39%
Signal transduction mechanisms	**T**	5.42%	6.92%	4.89%
Cell wall/membrane/envelope biogenesis	**M**	5.49%	5.60%	5.36%
Cell motility	**N**	4.31%	2.73%	1.51%
Cytoskeleton	**Z**	0.00%	0.01%	0.01%
Extracellular structures	**W**	0.00%	0.00%	0.00%
Intracellular trafficking, secretion, and vesicular transport	**U**	1.74%	1.44%	1.53%
Posttranslational modification, protein turnover, chaperones	**O**	3.13%	2.76%	2.99%
Metabolism				
Energy production and conversion	**C**	5.97%	5.94%	5.17%
Carbohydrate transport and metabolism	**G**	16.60%	7.26%	8.09%
Amino acid transport and metabolism	**E**	10.56%	8.42%	8.86%
Nucleotide transport and metabolism	**F**	2.78%	2.84%	3.33%
Coenzyme transport and metabolism	**H**	5.00%	4.13%	4.10%
Lipid transport and metabolism	**I**	2.22%	1.97%	2.36%
Inorganic ion transport and metabolism	**P**	5.28%	4.87%	5.22%
Secondary metabolites biosynthesis, transport and catabolism	**Q**	0.49%	1.03%	1.10%
Poorly characterized				
General function prediction only	**R**	11.46%	11.09%	11.20%
Function unknown	**S**	6.53%	8.36%	9.39%
Total Genomes Analyzed		1	32	300

category in *Epulopiscium* when compared to either the Firmicutes or *Clostridium*, with the exception of carbohydrate transport and metabolism, which had almost double the proportion. Taken together, these data indicate that this draft is somewhat incomplete, although it is difficult to say with absolute certainty the level of completeness.

For the identification of potential sporulation gene homologs in *Epulopiscium*, contigs and singletons were compiled into a pseudomolecule by linking all end-to-end. A spreadsheet was developed listing the position of each junction between adjacent fragments so that any chimeric open reading frames formed by during pseudomolecule assembly could be identified.

The core sporulation gene list. As a starting point, a comprehensive list of sporulation genes was compiled based on studies of *B. subtilis* 168 and its derivatives since this is by far the most thoroughly characterized sporulation program. The initial list of 732 genes included those described using classic genetic approaches as well as putative sporulation genes uncovered in transcriptional array analyses of sporulation regulons (30, 53, 65, 85, 94). BLAST searches refined the list, narrowing it to genes that are conserved in both endospore-forming Bacilli and Clostridia. Further refinement of the list eliminated genes with homologs in non-spore-forming Firmicutes.

Several notable genes that were eliminated are essential in the *B. subtilis* model of sporulation, including the phosphorelay protein Spo0B, and associated histidine kinases, transcription factors RsfA, GerR and GerE, and the intercellular signal transduction proteins SpoIIQ and SpoIVFA. The limited distribution of the phosphorelay among endospore-forming bacteria has been noted previously (87, 88) and it is likely that clostridia use a different system for activation of Spo0A (69, 86). In agreement with a previous study (26), we were unable to identify a SpoIIQ homolog in the clostridia.

A few key early sporulation proteins (Spo0A, Spo0J, Soj, SpoIIIE and SpoIIIJ), known to function during normal growth in *B. subtilis*, were placed back on the core list. All of these are part of the σ^A regulon and were first identified as essential for efficient sporulation. Spo0A determines entry into sporulation. Soj and Spo0J are members of the ParA/B family of

partitioning proteins that regulate transcription of early sporulation genes and assist in maintaining chromosome architecture (43). The FtsK homolog SpoIIIE aids in chromosome separation during binary fission (83), and during sporulation it is responsible for chromosome translocation into the developing forespore (12, 96). The location of genes on the chromosome and timing of translocation affects cell-specific expression of genes, which impacts key molecular events such as the activation of sporulation sigma factors (18, 31, 60). SpoIIIJ is a membrane protein translocase that, in addition to its vegetative growth function, appears to play a key role in the activation of σ^G during sporulation (34, 77). In the end, our core list included 147 genes (Table 2.3).

An overview of the core sporulation genes found in the *Epulopiscium* and *C. lentocellum*. We were concerned that the phylogenetic distance between *B. subtilis* and *Epulopiscium* may bias the distribution of conserved genes we would be able to identify. To gauge how much evolutionary divergence is impacting the recovery of homologs in *Epulopiscium* we compared our core list of sporulation genes to the *C. lentocellum* genome as well. *Cellulosilyticum lentocellum* DSM 5427 is the closest endospore-forming relative of *Epulopiscium* with an available completely sequenced genome (64). *Epulopiscium* sp. type B and *C. lentocellum* are members of the Lachnospiraceae and the two share 91% rRNA sequence identity. Of the 4185 predicted protein-coding genes in the *C. lentocellum* genome, an *Epulopiscium* sp. type B gene is the top BLAST hit for 546 of these genes (data not shown).

Of the 147 core sporulation genes, we found 87 homologs in the *C. lentocellum* genome and 57 putative homologs in the *Epulopiscium* sp. type B genome (Table 2.4). All of the sporulation genes found in the *Epulopiscium* genome were also found in the *C. lentocellum* genome. In the list of conserved genes, two code for members of a family of functionally

Table 2.3. Core sporulation genes

GENE	REGULON(S)	FUNCTION
aprX	K	alkaline serine protease
bofA	E	inhibitor of the pro-σ^K processing machinery
cotI	K	spore coat protein
cotJB	E	component of the inner spore coat
cotJC	E	component of the inner spore coat
cotS	K	spore coat protein
csfB	F	anti-σ^G factor
cwlD	E, G	N-acetylmuramoyl-L-alanine amidase/germination
cwlJ	E	cell wall hydrolase/germination, cortex lytic
dacB	E	D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein 5)
dacF	F, G	D-alanyl-D-alanine carboxypeptidase (penicilin-binding protein)
gerAA	F, G	component of the germination receptor GerA
gerAB	F, G	component of the germination receptor GerA
gerAC	F, G	component of the germination receptor GerA
gerBA	G	component of germinant receptor B
gerBB	G	component of germinant receptor B
gerBC	G	lipoprotein component of the germination receptor B
gerKA	G	spore germination receptor subunit
gerKB	G	spore germination receptor subunit
gerKC	G	spore germination receptor subunit
gpr	F, G	germination protease
jag	A	SpoIIIJ-associated RNA/ssDNA-binding protein
kamA	E	lysine 2,3-aminomutase
kapD	E,K	inhibitor of the KinA pathway to sporulation
lonB	F	ATP-dependent protease/forespore-specific
mmgC	E	short chain acyl-CoA dehydrogenase
ntdA	E	biosynthesis of neotrehalosadiamine (amino-sugar antibiotic)/aminotransferase
pbpG	F, G	penicillin-binding protein (also known as ywhe)
pbpI	E, F	penicillin-binding protein PBP4B/mother cell specific
pdaA	G	exported N-acetylmuramic acid deacetylase/cortex lysis
prkA	E	serine protein kinase/not well characterized
sigE	A, 0A	sporulation sigma factor/mother cell only
sigF	H, 0A	sporulation sigma factor/forespore only
sigG	F, G	sporulation sigma factor/forespore only
sigK	E, K	sporulation sigma factor/mother cell only
sleB	G	spore cortex-lytic enzyme
soj	A, 0A	chromosome partitioning protein/transcriptional regulator/negative regulation of sporulation initiation
splB	G	spore photoproduct (thymine dimer) lyase
spmA	E	spore maturation protein/spore dehydration

Table 2.3 (Continued)

GENE	REGULON(S)	FUNCTION
spmB	E	spore maturation protein/spore dehydration
spo0A	A, H, 0A	two-component response regulator central for the initiation of sporulation/"master regulator"
spo0F	H, 0A	two-component response regulator involved pathway leading to phosphorylation of Spo0A
spo0J	A	site-specific DNA-binding protein/chromosome positioning near the pole and transport through the polar septum /antagonist of Soj-dependent inhibition of sporulation initiation
spoIIAA	H, 0A	anti-anti-sigma factor (antagonist of SpoIIAB)
spoIIAB	H, 0A	anti-σ^F factor
spoIID	E	autolysin required for complete dissolution of the asymmetric septum
spoIIE	A, 0A	serine phosphatase (σ^F activation) /polar septum formation
spoIIGA	A, 0A	protease processing pro-σ^E
spoIIIAA	E	ATP-binding stage III sporulation protein/mother cell signalling for activation of σ^G
spoIIIAB	E	stage III sporulation protein/mother cell signalling for activation of σ^G
spoIIIAC	E	stage III sporulation protein/mother cell signalling for activation of σ^G
spoIIIAD	E	stage III sporulation protein/mother cell signalling for activation of σ^G
spoIIIAE	E	stage III sporulation protein/mother cell signalling for activation of σ^G
spoIIIAF	E	stage III sporulation protein/mother cell signalling for activation of σ^G
spoIIIAG	E	stage III sporulation engulfment assembly protein/mother cell signalling for activation of σ^G
spoIIIAH	E	stage III sporulation ratchet engulfment protein/mother cell signalling for activation of σ^G
spoIIID	E	transcriptional regulator of σ^E and σ^K-dependent genes
spoIIIE	A	spore DNA translocase
spoIIIJ	A	protein translocase/essential for activation of σ^G
spoIIM	E	autolysin component for dissolution of the septal cell wall
spoIIP	E, F, G	spore autolysin
spoIIR	F	endopeptidase/activation of σ^E
spoIVA	E	morphogenetic protein required for proper spore cortex formation and coat assembly
spoIVB	F, G, 0A	regulatory membrane-associated serine protease/intercompartmental signalling of pro-σ^K processing and activation in the mother-cell
spoIVFB	E	membrane metalloprotease required for the processing of pro-σ^K to active σ^K
spoVAA	G	essential for dipicolinic acid uptake by the developing spore
spoVAB	G	essential for dipicolinic acid uptake by the developing spore
spoVAC	G	essential for dipicolinic acid uptake by the developing spore
spoVAD	G	essential for dipicolinic acid uptake by the developing spore
spoVAEB	G	essential for dipicolinic acid uptake by the developing spore
spoVAF	G	essential for dipicolinic acid uptake by the developing spore

Table 2.3 (Continued)

GENE	REGULON(S)	FUNCTION
spoVB	E	spore cortex synthesis
spoVD	E	penicillin-binding protein
spoVE	E	factor for spore cortex peptidoglycan synthesis
spoVFA	K	spore dipicolinate synthase subunit A
spoVFB	K	spore dipicolinate synthase subunit B
spoVR	E	involved in spore cortex synthesis
spoVS	H	regulator required for dehydratation of the spore core and assembly of the coat
spoVT	F, G	transcriptional positive and negative regulator of σ^G-dependent genes
spsF	K	putative glycosyltransferase/spore coat polysaccharide synthesis
spsG	E, K	putative glycosyltransferase/spore coat polysaccharide synthesis
spsJ	E, K	putative dTDP-glucose 4,6-dehydratase/spore coat polysaccharide synthesis
sspA	G	small acid-soluble spore protein (alpha-type SASP)
sspB	G	small acid-soluble spore protein (beta-type SASP)
sspC	G	small acid-soluble spore protein (alpha/beta-type SASP)/SPβ phage protein
sspD	G	small acid-soluble spore protein (alpha/beta-type SASP)
sspF	G	small acid-soluble spore protein (alpha/beta-type SASP)
tlp	F, G	small acid-soluble spore protein
yaaH	E	spore peptidoglycan hydrolase
yabG	K	sporulation-specific protease
yabP	E	spore protein involved in the shaping of the spore coat
yabQ	E	membrane protein of the forespore/essential for spore cortex
ybaN	E	polysaccharide deacetylase involved in sporulation
ybdM	G	putative protein kinase
ycgF	E	putative aminoacid export permease
ydfS	K, G	hypothetical protein
ydhD	E	spore cortex lytic enzyme
yerB	0A	putative lipoprotein
yfkQ	G	putative spore germination protein
yfkR	G	putative spore germination protein
yfkT	G	putative spore germination integral inner membrane protein
yfnG	K	putative sugar-phosphate cytidylyltransferase
yfnH	K	putative FAD-dependent oxido-reductase
ygaK	K	putative FAD-dependent oxido-reductase
yhbH	E	hypothetical protein
yhcB	G	putative oxidoreductase associated to oxygen stress
yhcV	G	putative oxidoreductase
yhfW	F	putative Rieske [2Fe-2S] oxygenase
yisY	E	putative hydrolase

Table 2.3 (Continued)

GENE	REGULON(S)	FUNCTION
ykuD	K	murein transglycosylase
ykuS	F	hypothetical protein
ykvU	E	spore membrane protein involved in germination
ylaK	E	putative phosphate starvation inducible protein
ylbB	F	putative oxidoreductase
ylbJ	E	putative factor required for spore cortex formation
ymxH	E	hypothetical protein
yndD	G	putative spore germination protein
yndE	G	putative spore germination integral inner membrane protein
yndF	G	putative spore germination lipoprotein
yngE	E	putative propionyl-CoA carboxylase
yngI	E	AMP-binding domain protein
yngJ	E	acyl-CoA dehydrogenase, short-chain specific
yoaR	G	putative factor for cell wall maintenance or synthesis
yobN	E	putative amine oxidase
ypeB	G	spore membrane component
yqfC	E	hypothetical protein
yqfD	E	stage IV sporulation protein
yqgT	E	putative gamma-D-glutamyl-L-diamino acid endopeptidase
yqhO	E	hypothetical protein
yraD	G	putative spore coat protein
yrbG	E, G	hypothetical protein
yrkC	K	putative dioxygenase; cupin family
ytcA	K	putative UDP-glucose dehydrogenase
ytcC	K	putative glucosyltransferase
ytfJ	F	hypothetical protein
ytlA	K	putative ABC transporter component
ytlC	K	putative ABC transporter component, ATP-binding
ytlD	K	putative permease of ABC transporter
ytvI	E	putative permease
ytxC	E	hypothetical protein
yunB	E, K	putative protein involved in spore formation
yutH	F	spore coat-associated protein
ywjD	G	putative UV damage endonuclease
yyaC	F	hypothetical protein
yyaE	E	putative oxidoreductase
yyaO	K	hypothetical protein
yyBI	E	inner spore coat protein

Table 2.4. Genes conserved in *Epulopiscium* and *C. lentocellum*

GENE	PROMOTER(S)	In *Epulopiscium*	In *C. lentocellum*
Genes in the σ^A regulon			
jag			X
sigE		X	X
soj		X	X
spo0A		X	X
spo0J		X	X
spoIIE		X	X
spoIIGA		X	X
spoIIIE	σ^A	X	X
spoIIIJ		X	X
Genes in the σ^H regulon			
sigF		X	X
spoIIAA		X	X
spoIIAB		X	X
Genes in the σ^F regulon			
dacF	σ^G	X	X
gpr	σ^G	X	X
lonB		X	X
sigG	σ^F, σ^G	X	X
spoIIR		X	X
spoIVB	σ^F, σ^G	X	X
spoVT	σ^G	X	X
yhfW			X
ytfJ			X
yyaC		X	X
Genes in the σ^E regulon			
bofA			X
cotJB			X
cotJC			X
cwlD	σ^G	X	X
dacB		X	X
kamA			X
prkA		X	X
sigK		X	X
spmA		X	X
spmB		X	X
spoIID	σ^E	X	X
spoIIIAA	σ^E	X	X

Table 2.4 (Continued)

GENE	PROMOTER(S)	In *Epulopiscium*	In *C. lentocellum*
spoIIIAB	σ^E	**X**	**X**
spoIIIAC	σ^E	**X**	**X**
spoIIIAD	σ^E	**X**	**X**
spoIIIAE	σ^E	**X**	**X**
spoIIIAF	σ^E	**X**	**X**
spoIIIAG	σ^E	**X**	**X**
spoIIIAH	σ^E	**X**	**X**
spoIIID	σ^E	**X**	**X**
spoIIP	σ^G	**X**	**X**
spoIVA	σ^E	**X**	**X**
spoVB		**X**	**X**
spoVD		**X**	**X**
spoVE		**X**	**X**
spoVR		**X**	**X**
spsJ			**X**
yaaH			**X**
yabP	σ^E	**X**	**X**
yabQ	σ^E	**X**	**X**
yhbH		**X**	**X**
ylbJ			**X**
yngI			**X**
yqfC	σ^E	**X**	**X**
yqfD	σ^E	**X**	**X**
yqgT			**X**
yqhO			**X**
ytvI		**X**	**X**
Genes in the σ^G regulon			
gerKA			**X**
gerKB			**X**
gerKC			**X**
pdaA		**X**	**X**
splB			**X**
spoVAA			**X**
spoVAB			**X**
spoVAC			**X**
spoVAD			**X**
spoVAEB			**X**
spoVAF		**X**	**X**

Table 2.4 (Continued)

GENE	PROMOTER(S)	In *Epulopiscium*	In *C. lentocellum*
sspB		**X**	**X**
sspC/F		**X**	**X**
ybdM		**X**	**X**
ydfS			**X**
Genes in the σ^K regulon			
cotS		**X**	**X**
spoVFA			**X**
spoVFB	σ^K	**X**	**X**
yabG		**X**	**X**
yfnG		**X**	**X**
yfnH		**X**	**X**
ygaK			**X**
yrkC			**X**
ytcC			**X**
ytlA			**X**
ytlC			**X**
ytlD			**X**

redundant small acid-soluble proteins (SASPs) that bind to and protect spore DNA from damage (78, 80). One of the SASPs is similar to two paralogs of this family from *B. subtilis* and is represented as *sspC/F*. In both *Epulopiscium* and *C. lentocellum*, at least five genes are represented from each of the four sporulation sigma factor regulons, as well as genes in the Spo0A regulon with σ^H and σ^A promoters. Of the *C. lentocellum* homologs not seen in the *Epulopiscium* genome, 14 code for "y" genes that are expressed during sporulation, although their roles have yet to be characterized. These genes are equally distributed between pre- and post-engulfment regulons, but only two are expressed in the forespore in *B. subtilis*.

Conservation of the master regulator Spo0A in *Epulopiscium*. Although clostridia do not have a phosphorelay system homologous to that used by *B. subtilis*, Spo0A is conserved in the genomes of all known endospore-forming bacteria. The *Epulopiscium* sp. type B genome contains an unambiguous Spo0A homolog. This suggests that although offspring production seems hard-wired in *Epulopiscium* and there is no obvious need to adapt the timing of entry into offspring production to environmental conditions, Spo0A is still important for initiating this process or possibly other metabolic transitions. *Bacillus* spp. use the phosphorylation state and abundance of Spo0A to engage in activities, such as cannibalism, to maintain a growing population for as long as possible prior to resorting to sporulation (38). In clostridia, Spo0A also dictates metabolic transitions. For example, in *Clostridium acetobutylicum*, Spo0A activation shifts the cell from acid production during exponential phase growth to solvent production during stationary phase (73). The identification and characterization of kinases or phosphatases in *Epulopiscium* that interact with Spo0A may provide information about the environmental and cellular cues that regulate population level developmental or physiological transitions.

The central regulatory network conserved in endospore-forming bacteria appears conserved in *Epulopiscium*. In *B. subtilis*, the sequential activation of cell-specific sigma factors directs compartmentalized gene expression that determines the different fates of the mother cell and forespore (26). The presence of homologous genes suggests a similar system likely also functions in *Epulopiscium* developmental progression (Figure 2.3). In a previous publication, we described the σ^F homolog of *Epulopiscium* and a structural analysis of the proteins required for σ^F activation: SpoIIE, SpoIIAA, and SpoIIAB (63). In the present study, we uncovered all additional sporulation sigma factors (σ^E, σ^G and σ^K) and homologs for the intercellular signal transduction proteins responsible for the activation of σ^E (SpoIIR and SpoIIGA) and σ^G (SpoIIIAA-AH and SpoIIIJ). In *B. subtilis* SpoIIQ interacts with SpoIIIAA-AH in an intracellular signaling system required for the activation of σ^G (61). The gene coding for SpoIIQ is absent in clostridia (26) therefore involvement of SpoIIQ is a *Bacillus*-specific innovation that likely does not represent the ancestral mode of cell-cell communication used to trigger late forespore sigma activation.

Mechanisms to activate the late mother-cell sigma, σ^K, are apparently the most specialized of the sporulation sigma factors. In *B. subtilis*, the gene encoding σ^K contains a 48 kb insertion sequence called the *skin* (sigma K intervening) element that must first be excised from the mother-cell chromosome by the recombinase SpoIVCA (72, 89). The reconstituted gene produces an inactive pro-protein which must be processed by SpoIVFB but BofA and SpoIVFA inhibit SpoIVFB activity (24, 74). These three proteins are transcribed by σ^E in the mother cell and localize to the outer forespore membrane. SpoIVB and CtpB, regulated by σ^G in the forespore, inhibit the actions of BofA and SpoIVFA allowing cleavage of pro-σ^K (17, 23). In this way, σ^K operates in the mother cell only after σ^G has been activated in the forespore. Only

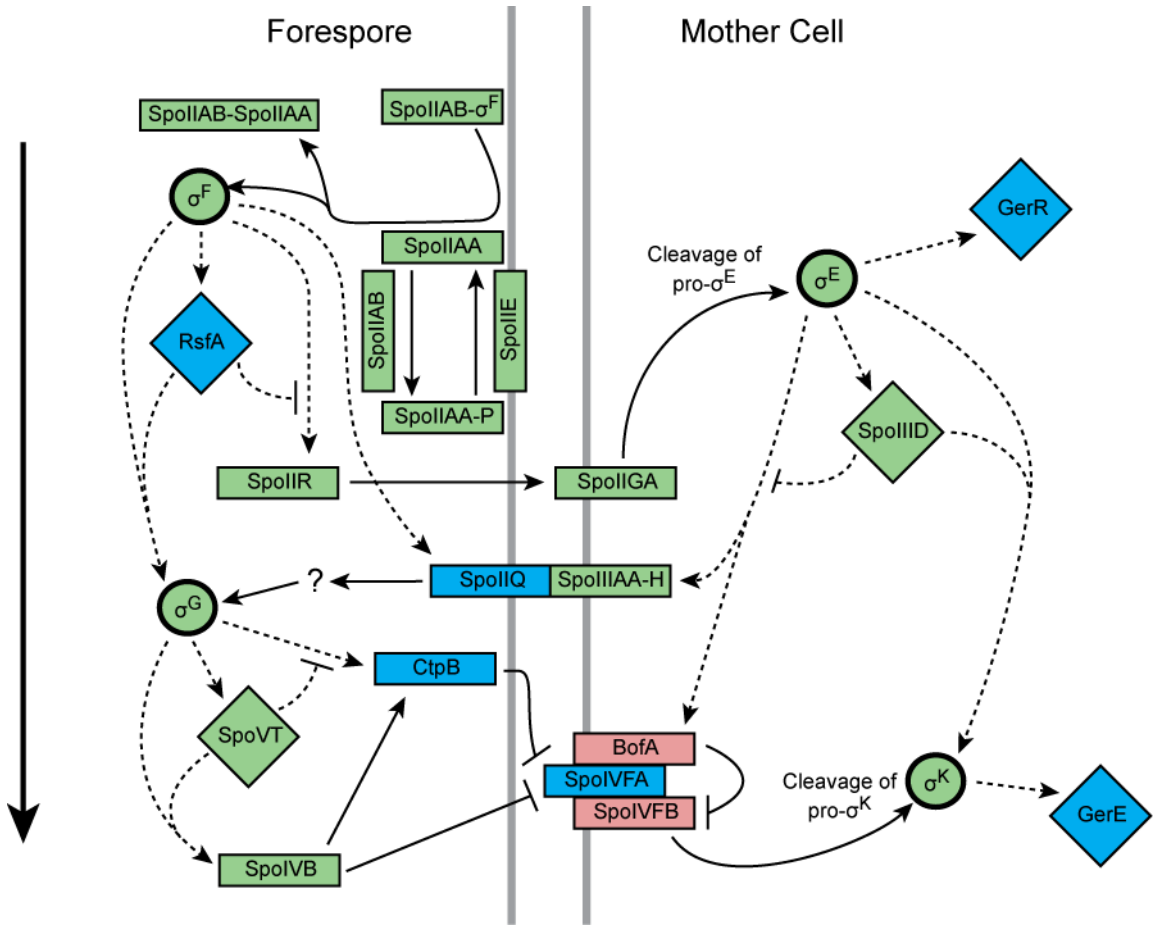


Figure 2.3. Conservation of the *B. subtilis* sporulation regulatory cascade in *Epulopiscium* sp. type B. Sporulation-specific sigma factors (circles), associated transcription factors (diamonds) and other signal transduction or regulatory proteins involved in sigma activation (rectangles) are shown. Colors of the proteins indicate the gene presence in *Epulopiscium* (green), on the core list but not in *Epulopiscium* (red), and absence from the core list (blue). Control of gene expression is indicated by dotted lines and arrows. Signaling pathways and other protein interactions are denoted with solid lines and arrows. Temporal transcriptional progression through the cascade is shown by the position on the diagram with earlier stages near the top. A detailed explanation of the regulatory cascade as it occurs in *B. subtilis* is provided in the text. Figure is adapted from de Hoon *et al*. (2010).

homologs for *spoIVB* and *ctpB* were found in *Epulopiscium*. Note that *ctpB* homologs are found in many non-spore-forming bacteria, which is why it is not included in the core list. The gene coding for SpoIVFA is absent in clostridia and although *spoIVFB* and *bofA* met the requirements to be included on the core list, neither is conserved in many clostridia and only *bofA* was found in the *C. lentocellum* genome. The weak conservation of these proteins outside of the Bacilli indicates that the intricate mechanism of σ^K activation described in *B. subtilis* is a recent innovation.

In addition to conserved sporulation sigma factors and many of the proteins responsible for regulating their activation, we found homologs of transcription factors SpoIIID and SpoVT. These proteins work together with their associated sigma factor (σ^E and σ^G, respectively) to modulate transcription of genes downstream in the process. Other transcription factors that function during sporulation in *B. subtilis* (RsfA, GerR and GerE) are not conserved in the clostridia. These feed-forward loops appear to be a Bacilli-specific amendment. Based on these results, we conclude that gene regulation by sporulation-specific sigma factors and their associated transcription factors involved in their sequential activation of cell-specific genetic programs is highly conserved in *Epulopiscium*. With the exception of *bofA*, all of the core sporulation genes coding for components of the central regulatory cascade that are conserved in the *C. lentocellum* genome are also conserved in *Epulopiscium*.

Engulfment genes are highly conserved in *Epulopiscium*. SpoIID, SpoIIP and SpoIIM are required for degradation of septal peptidoglycan and progression toward forespore engulfment in *B. subtilis* (20). We found *Epulopiscium* genes that code for homologs of SpoIID and SpoIIP but could not identify a SpoIIM homolog. SpoIIM appears highly conserved in spore-forming bacteria but is curiously absent from the Lachnospiraceae lineage, as the genome of

C. lentocellum also lacks an apparent homolog of *spoIIM* (Table 2.4). The *Epulopiscium spoIIP* initially retrieved from the pseudomolecule did not meet the requirements to be considered a homolog in this analysis, as it appeared to be missing its 5´ end. In *B. subtilis*, *spoIIP* has its own σ^E promoter (37) but is also the second gene in an operon with *gpr* (94), and this operon structure is widely conserved in *Bacillus* and *Clostridium* spp. A truncated *gpr* homolog was identified in the *Epulopiscium* draft genome. Based on this information, PCR and sequence analysis was used to recover the complete *Epulopiscium gpr - spoIIP* operon.

Other proteins that function during engulfment in *B. subtilis* are SpoIIB, SpoIIIE, SpoIIIAH and SpoIIQ. SpoIIB likely plays a role in the localization of the SpoIIDMP complex in *B. subtilis* (10) but it is not conserved in the clostridia. Null mutants of *spoIIB* in *B. subtilis* are oligosporogenous, and this sporulation defect becomes more pronounced when combined with a *spoVG* mutation (59). In addition to its DNA translocase activity, SpoIIIE has been implicated in the fusion of the leading edge of the mother-cell membrane after migrating around the forespore (54, 82). We identified homologs of SpoIIIE in both *Epulopiscium* and *C. lentocellum.* Beyond the role of the SpoIIIAH - SpoIIQ complex in σ^G activation, these two proteins may perform an important role during engulfment. It has been suggested that as the mother-cell membrane migrates around the forespore, these proteins bind to one another near the leading edge of the mother-cell membrane to prevent it from retreating back toward the midcell (13). Homologs of the *spoIIIA* operon were identified in *C. lentocellum* and *Epulopiscium*. With the exception of SpoIIM and SpoIIQ, all of the known proteins required for engulfment in *B. subtilis* are coded for in the *Epulopiscium* genome (Figure 2.4). Neither SpoIIM nor SpoIIQ are coded for in the *C. lentocellum* genome.

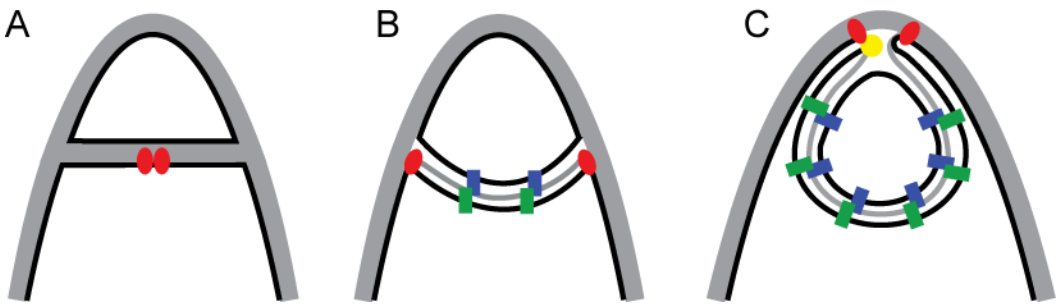


Figure 2.4. *Epulopiscium* engulfment model. The *Epulopiscium* genome codes for all of the genes known to be essential for engulfment in *B. subtilis*, except *spoIIM* and *spoIIQ* which are also absent in *C. lentocellum*. A) SpoIID and SpoIIP assemble into a complex (red ovals) at the division septum and degrade the septal peptidoglycan. B) As the mother cell membrane wraps around the offspring, the IIDP complex tracks along the leading edge where it is involved in interactions with the mother-cell peptidoglycan and synthesis of offspring cell wall. C) When it reaches the cell tip, membrane fusion is mediated by SpoIIIE (yellow circle). During engulfment, SpoIIIAH (green rectangles) produced in the mother cell and a hypothetical protein (blue rectangles) from the offspring cell bind and prevent backward movement of the mother-cell membrane. In this diagram, black lines indicate membranes and grey peptidoglycan.

Late stage sporulation proteins in the *Epulopiscium* genome. A small number of late sporulation genes were identified in the *Epulopiscium* sp. type B genome. This may be an underestimate as many of the proteins in this category are very small and functionally redundant. For instance, it is possible that some coat proteins have diverged to a degree that they could not to be detected by our methods. Regardless, *B. subtilis* uses more than 75 proteins for cortex and coat formation, spore maturation and germination of which *Epulopiscium* has retained 17 possible homologs. For comparison, the *C. lentocellum* genome has 31. The genes retained in

Epulopiscium may perform novel functions or they are relics of its spore-forming ancestry. Previous studies have identified sporulation homologs in the genomes of non-sporulating members of the Firmicutes (68). Additionally, many surgeonfish intestinal symbionts that are close relatives of *Epulopiscium* sp. type B form endospores (36) and would likely need late sporulation proteins for the formation of cortex and coat as well as spore maturation and germination similar to those in *B. subtilis* and the clostridia.

Based on their functions in *B. subtilis*, we classified the late sporulation genes located in *C. lentocellum* and/or *Epulopiscium*. With the notable exception of genes coding for SASPs (*sspB*, *sspC/F*) or SASP degradation (*gpr*), genes involved in resistance properties of a mature spore and germination signal receptors (*gerK* operon) are more highly conserved in *C. lentocellum*. Specifically, *C. lentocellum* has the nine genes required for the synthesis and forespore-uptake of dipicolinic acid (DPA) while *Epulopiscium* has only one of these genes. Likewise, only *C. lentocellum* has *splB*, which codes for the DNA repair enzyme spore-photoproduct lyase. Surprisingly, all of the ten genes on the list involved in cortex biosynthesis or other cortex-associated properties that are conserved in *C. lentocellum* are also found in *Epulopiscium*. In addition, *spoIVA*, *yabP* and *yabQ*, which encode scaffolding proteins important for cortex and coat morphogenesis (9, 71, 92), were found in both *C. lentocellum* and *Epulopiscium*. Four coat protein genes were found in *C. lentocellum* but only one of these appears conserved in *Epulopiscium*. The biased distribution of functional categories represented in *Epulopiscium* suggests that cortex biosynthetic machinery has been retained, as well as the DNA-protective SASPs, and these proteins may still serve some function. Overall, the pattern of conserved genes in *C. lentocellum* compared with *Epulopiscium* (Figure 2.5) supports our hypothesis that there is a strong selection to retain early stage sporulation gene homologs in

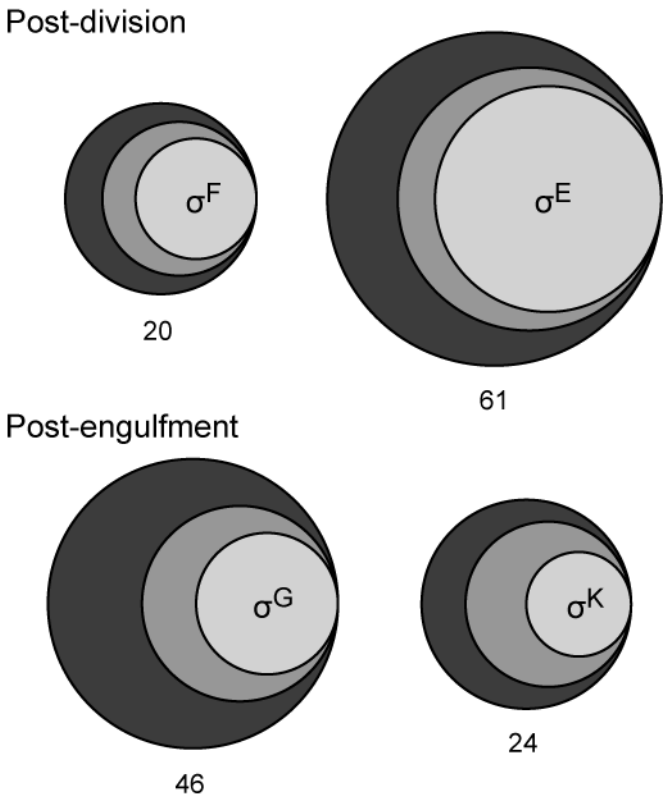


Figure 5. Distribution of core sporulation genes conserved in the *Epulopiscium* and *C. lentocellum* genomes by regulon. Venn diagrams represent the conservation of genes in the four sporulation-specific sigma factor regulons. Circle size corresponds to the number genes on the core list for *B. subtilis* (outermost circle), *C. lentocellum* (middle circle) and *Epulopiscium* (inner circle). The numbers below each diagram indicate the total number of genes from *B. subtilis* in each regulon. Some genes were counted more than once if they are members of multiple regulons as indicated in Table 2.3.

Epulopiscium such as those necessary for compartmentalized gene expression, engulfment and developmental progression.

Promoter analysis of putative developmentally regulated genes. For each of the sporulation genes found in the *Epulopiscium* genome, the region immediately upstream of the predicted start

Table 5. Promoters found upstream of sporulation gene homologs in the *Epulopiscium* genome

GENE/OPERON	PROMOTER							
	σ^F		σ^E		σ^G		σ^K	
	-35	-10	-35	-10	-35	-10	-35	-10
sigG	GTATA	GGGTATCCTA			GTATA	TATCCTA		
spoIVB	GTATA	GGCAATTTTA			GTATA	AATTTTA		
spoIID			TTATAAGT	TCATATAATT				
spoIIIA			TAATGATT	GCATATACTG				
spoIIID			TAATATAT	GCATATTATT				
spoIVA			TCATATCC	ACATATAGTT				
yabPQ			GAATAATT	AAATATAAAT				
yqfCD			GAATACTT	GCATAATATG				
dacF					GTATA	AATAATA		
cwlD					GAATT	GATAATA		
gpr-spoIIP					GATTA	TATATTA		
spoVT					GCATA	CATAATA		
spoVFB							TCACA	TCATATTATA
B. subtilis consensus[a]	**G**TA**TA**	**G**GNA**A**N**A**M**T**R	TYA**T**ATTT	KC**ATA**NAN**T**N	GN**AT**A	C**A**W**A**M**TA**	KC**AC**M	GC**ATA**NNN**T**A

[a] Consensus sequences from *B. subtilis* were based on Wang *et al*. (2006) and Eichenberger *et al.* (2004). Bold indicates a highly conserved base. N = A, T, G, C; R = A, G; Y = C, T; K = T, G; M = A, C; W = A, T.

codon (~300 bp) was visually scanned for potential promoter sequences (Table 2.5). While this analysis is not definitive, we reasoned that identifiable promoters that match their predicted *B. subtilis* counterparts would reinforce the predictions placing genes in particular regulons. Consensus of the -10 and -35 promoter sequences from *B. subtilis* promoters were used (3, 30, 94). We were able to identify σ^F promoters upstream of *sigG* and *spoIVB*, and σ^E promoters upstream of *spoIID*, the *spoIIIA* operon, *spoIIID*, the *yabPQ* operon and the *yqfCD* operon. σ^G promoters were found upstream of *dacF*, *cwlD*, *sigG*, *spoIVB*, *spoVT* and the *gpr – spoIIP* operon. Only one gene, *spoVFB*, was found to have a σ^K promoter. Our analysis of the putative *dacF* homolog indicated that approximately 50% of the 5´ end of the gene was present in the pseudomolecule, however, we were unable to obtain a full-length gene by PCR. The gene *spoVFB* codes for the ß-subunit of DPA synthase and is usually located in a bicistonic operon downstream of *spoVFA*, the gene for the α-subunit of the DPA synthase (25). DPA is found in abundance only in mature endospores (33) and is a distinct feature of the phase-bright spores of the *Epulopiscium*-like symbionts of *Naso lituratus* (36). Although we were unable to recover *spoVFA* from *Epulopiscium*, we did find homologs of *spoVFA* and *spoVFB* adjacent to each other in the *C. lentocellum* genome.

Conclusions. The comparative analysis of the draft *Epulopiscium* sp. type B genome with the complete *C. lentocellum* genome substantiates our hypothesis that the production of intracellular offspring in *Epulopiscium* evolved from endospore formation. All of the genes identified in *C. lentocellum* that function in engulfment as well as the core transcriptional regulatory cascade, and the associated intracellular communication network that coordinates sigma factor activation, were found in the *Epulopiscium* genome. While we could identify homologs of late sporulation genes, a large proportion of the late sporulation genes found in *C. lentocellum* was not recovered

from *Epulopiscium*. Since we used a draft genome to explore the conservation of sporulation genes in *Epulopiscium*, it is possible that we did not recover all of the sporulation genes retained in this genome, however, we would not expect a functional bias in the distribution of genes recovered. Therefore the fewer late genes recovered in *Epulopiscium* probably reflect the decline of this class of genes in evolution.

Although *Epulopiscium* sp. type B is closely related to surgeonfish intestinal symbionts that form multiple endospores to reproduce, it appears that type B cells may no longer have the genetic capacity to form a dormant and fully resistant endospore. Many intestinal anaerobes, from harmful pathogens to benign commensals, use endospores for effective dispersal between vertebrate hosts (2, 7, 19). The ability of an intestinal bacterium to produce an endospore should be valuable to survival. What selection drove *Epulopiscium* sp. type B to give up this trait? We speculate that the perpetuation of large cell size and the ability to maintain a longer residence in an individual host may be factors that contributed to the loss of dormancy and associated resistance traits in offspring development in this lineage. The large size of *Epulopiscium* sp. type B cells may be important to maintain their position in the gut and to avoid predation by the ciliate predators that cohabitate the *N. tonganus* intestinal tract (62). The growth and development of offspring within a live mother cell may be essential for maintaining these benefits throughout the life cycle of an individual. It is also possible that the manifestation of some resistance traits may simply be impossible for a cell as large as an *Epulopiscium* sp. type B offspring. For example, the formation of a flawless spore coat may be physically impossible for this size of cell.

The study outlined here presents the first genome sequence for any *Epulopiscium* species and further provides insight into the evolution of a novel form of cellular reproduction in the

Epulopiscium lineage. Based on its phylogenetic position among endospore-forming lineages, we reason that formation of active intracellular offspring in *Epulopiscium* sp. type B is a recent modification of the sporulation program. Our comparative study reveals those essential physical (e.g. engulfment of the offspring cell) and regulatory mechanisms (e.g. alternative sigma factors and intracellular communication) that have been maintained in a live-offspring-bearing cell. Additionally, we suggest that those genes conserved in *Epulopiscium* sp. type B identify a subset of sporulation genes that may have been acquired early in the evolution of the Firmicutes that provided a selectable advantage to the predecessor of endospore-forming bacteria. For example, *Epulopiscium* has retained genes coding for SASPs and a protease (*gpr*) responsible for their degradation. These non-specific DNA binding proteins protect the spore DNA from damage by radiation, chemicals like hydrogen peroxide and heat (79, 81). Having SASPs could be advantageous to cells that live in fluctuating environments. The only other late sporulation genes that appear to be retained in *Epulopiscium* are those involved in cortex biosynthesis and function. The ability to synthesize a cortex-like peptidoglycan barrier, that would be less rigid than the typical cell-wall peptidoglycan of a Gram-positive Firmicutes, could be a distinct advantage for the rapid growth and development of internal offspring cells. It is also possible that the subset of resistance traits retained by *Epulopiscium* improve offspring survival during transit to a new host or during brief periods of nutrient deprivation.

Clearly further functional studies are required to test the above models and identify additional mechanisms acquired or modified during evolution of offspring formation in *Epulopiscium*. The findings presented here will provide a foundation to begin to assess genetic programs expressed during development in *Epulopiscium*.

REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389-402.
2. **Alvarez, Z., and E. Abel-Santos.** 2007. Potential use of inhibitors of bacteria spore germination in the prophylactic treatment of anthrax and *Clostridium difficile*-associated disease. Expert Rev. Anti. Infect. Ther. **5:**783-92.
3. **Amaya, E., A. Khvorova, and P. J. Piggot.** 2001. Analysis of promoter recognition *in vivo* directed by σ^F of *Bacillus subtilis* by using random-sequence oligonucleotides. J. Bacteriol. **183:**3623-30.
4. **Angert, E. R.** 2005. Alternatives to binary fission in bacteria. Nat. Rev. Microbiol. **3:**214-24.
5. **Angert, E. R., A. E. Brooks, and N. R. Pace.** 1996. Phylogenetic analysis of *Metabacterium polyspora*: clues to the evolutionary origin of daughter cell production in *Epulopiscium* species, the largest bacteria. J. Bacteriol. **178:**1451-6.
6. **Angert, E. R., and K. D. Clements.** 2004. Initiation of intracellular offspring in *Epulopiscium*. Mol. Microbiol. **51:**827-35.
7. **Angert, E. R., and R. M. Losick.** 1998. Propagation by sporulation in the guinea pig symbiont *Metabacterium polyspora*. Proc. Natl. Acad. Sci. U. S. A. **95:**10218-23.
8. **Arigoni, F., L. Duncan, S. Alper, R. Losick, and P. Stragier.** 1996. SpoIIE governs the phosphorylation state of a protein regulating transcription factor σ^F during sporulation in *Bacillus subtilis*. Proc. Natl. Acad. Sci. U. S. A. **93:**3238-42.

9. **Asai, K., H. Takamatsu, M. Iwano, T. Kodama, K. Watabe, and N. Ogasawara.** 2001. The *Bacillus subtilis yabQ* gene is essential for formation of the spore cortex. Microbiology **147:**919-27.

10. **Aung, S., J. Shum, A. Abanes-De Mello, D. H. Broder, J. Fredlund-Gutierrez, S. Chiba, and K. Pogliano.** 2007. Dual localization pathways for the engulfment proteins during *Bacillus subtilis* sporulation. Mol. Microbiol. **65:**1534-46.

11. **Bagyan, I., J. Hobot, and S. Cutting.** 1996. A compartmentalized regulator of developmental gene expression in *Bacillus subtilis*. J. Bacteriol. **178:**4500-7.

12. **Bath, J., L. J. Wu, J. Errington, and J. C. Wang.** 2000. Role of *Bacillus subtilis* SpoIIIE in DNA transport across the mother cell-prespore division septum. Science **290:**995-7.

13. **Broder, D. H., and K. Pogliano.** 2006. Forespore engulfment mediated by a ratchet-like mechanism. Cell **126:**917-28.

14. **Burbulys, D., K. A. Trach, and J. A. Hoch.** 1991. Initiation of sporulation in *B. subtilis* is controlled by a multicomponent phosphorelay. Cell **64:**545-52.

15. **Bylund, J. E., M. A. Haines, P. J. Piggot, and M. L. Higgins.** 1993. Axial filament formation in *Bacillus subtilis*: induction of nucleoids of increasing length after addition of chloramphenicol to exponential-phase cultures approaching stationary phase. J. Bacteriol. **175:**1886-90.

16. **Camp, A. H., and R. Losick.** 2008. A novel pathway of intercellular signalling in *Bacillus subtilis* involves a protein with similarity to a component of type III secretion channels. Mol. Microbiol. **69:**402-17.

17. **Campo, N., and D. Z. Rudner.** 2007. SpoIVB and CtpB are both forespore signals in the activation of the sporulation transcription factor σ^K in *Bacillus subtilis*. J. Bacteriol. **189:**6021-7.

18. **Chary, V. K., P. Xenopoulos, and P. J. Piggot.** 2006. Blocking chromosome translocation during sporulation of *Bacillus subtilis* can result in prespore-specific activation of σ^G that is independent of σ^E and of engulfment. J. Bacteriol. **188:**7267-73.

19. **Chase, D. G., and S. L. Erlandsen.** 1976. Evidence for a complex life cycle and endospore formation in the attached, filamentous, segmented bacterium from murine ileum. J. Bacteriol. **127:**572-83.

20. **Chastanet, A., and R. Losick.** 2007. Engulfment during sporulation in *Bacillus subtilis* is governed by a multi-protein complex containing tandemly acting autolysins. Mol. Microbiol. **64:**139-52.

21. **Clarkson, J., I. D. Campbell, and M. D. Yudkin.** 2004. Physical evidence for the induced release of the *Bacillus subtilis* transcription factor, σ^F , from its inhibitory complex. J. Mol. Biol. **340:**203-9.

22. **Clements, K. D., D. C. Sutton, and J. H. Choat.** 1989. Occurrence and characteristics of unusual protistan symbionts from surgeonfishes (Acanthuridae) of the Great Barrier Reef, Australia. Marine Biol. **102:**403-12.

23. **Cutting, S., A. Driks, R. Schmidt, B. Kunkel, and R. Losick.** 1991. Forespore-specific transcription of a gene in the signal transduction pathway that governs Pro-σ^K processing in *Bacillus subtilis*. Genes Dev. **5:**456-66.

24. **Cutting, S., S. Roels, and R. Losick.** 1991. Sporulation operon *spoIVF* and the characterization of mutations that uncouple mother-cell from forespore gene expression in *Bacillus subtilis*. J. Mol. Biol. **221:**1237-56.

25. **Daniel, R. A., and J. Errington.** 1993. Cloning, DNA sequence, functional analysis and transcriptional regulation of the genes encoding dipicolinic acid synthetase required for sporulation in *Bacillus subtilis*. J. Mol. Biol. **232:**468-83.

26. **de Hoon, M. J., P. Eichenberger, and D. Vitkup.** 2010. Hierarchical evolution of the bacterial sporulation network. Curr. Biol. **20:**R735-45.

27. **Delcher, A. L., K. A. Bratke, E. C. Powers, and S. L. Salzberg.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23:**673-9.

28. **Duncan, L., S. Alper, and R. Losick.** 1996. SpoIIAA governs the release of the cell-type specific transcription factor σ^F from its anti-sigma factor SpoIIAB. J. Mol. Biol. **260:**147-64.

29. **Duncan, L., and R. Losick.** 1993. SpoIIAB is an anti-sigma factor that binds to and inhibits transcription by regulatory protein σ^F from *Bacillus subtilis*. Proc. Natl. Acad. Sci. U. S. A. **90:**2325-9.

30. **Eichenberger, P., M. Fujita, S. T. Jensen, E. M. Conlon, D. Z. Rudner, S. T. Wang, C. Ferguson, K. Haga, T. Sato, J. S. Liu, and R. Losick.** 2004. The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. PLoS Biol. **2:**e328.

31. **Eldar, A., V. K. Chary, P. Xenopoulos, M. E. Fontes, O. C. Loson, J. Dworkin, P. J. Piggot, and M. B. Elowitz.** 2009. Partial penetrance facilitates developmental evolution in bacteria. Nature **460:**510-4.

32. **Ellar, D. J.** 1978. Spore specific structures and their function. Symp. Soc. Gen. Microbiol. **28:**295-325.

33. **Errington, J.** 1993. *Bacillus subtilis* sporulation: regulation of gene expression and control of morphogenesis. Microbiol. Rev. **57:**1-33.

34. **Errington, J., L. Appleby, R. A. Daniel, H. Goodfellow, S. R. Partridge, and M. D. Yudkin.** 1992. Structure and function of the *spoIIIJ* gene of *Bacillus subtilis*: a vegetatively expressed gene that is essential for σ^G activity at an intermediate stage of sporulation. J. Gen. Microbiol. **138:**2609-18.

35. **Fitz-James, P. C., and I. E. Young.** 1969. Morphology of sporulation, p. 39-72. *In* G. W. Gould and A. Hurst (ed.), The bacterial spore. Academic Press, New York, N.Y.

36. **Flint, J. F., D. Drzymalski, W. L. Montgomery, G. Southam, and E. R. Angert.** 2005. Nocturnal production of endospores in natural populations of *Epulopiscium*-like surgeonfish symbionts. J. Bacteriol. **187:**7460-70.

37. **Frandsen, N., and P. Stragier.** 1995. Identification and characterization of the *Bacillus subtilis spoIIP* locus. J. Bacteriol. **177:**716-22.

38. **Fujita, M., and R. Losick.** 2005. Evidence that entry into sporulation in *Bacillus subtilis* is governed by a gradual increase in the level and activity of the master regulator Spo0A. Genes Dev. **19:**2236-44.

39. **Gholamhoseinian, A., and P. J. Piggot.** 1989. Timing of *spoII* gene expression relative to septum formation during sporulation of *Bacillus subtilis*. J. Bacteriol. **171:**5747-9.

40. **Guzman, P., J. Westpheling, and P. Youngman.** 1988. Characterization of the promoter region of the *Bacillus subtilis spoIIE* operon. J. Bacteriol. **170:**1598-609.

41. **Halberg, R., and L. Kroos.** 1994. Sporulation regulatory protein SpoIIID from *Bacillus subtilis* activates and represses transcription by both mother-cell-specific forms of RNA polymerase. J. Mol. Biol. **243:**425-36.

42. **Henriques, A. O., and C. P. Moran, Jr.** 2000. Structure and assembly of the bacterial endospore coat. Methods **20:**95-110.

43. **Hilbert, D. W., and P. J. Piggot.** 2004. Compartmentalization of gene expression during *Bacillus subtilis* spore formation. Microbiol. Mol. Biol. Rev. **68:**234-62.

44. **Hofmeister, A. E., A. Londono-Vallejo, E. Harry, P. Stragier, and R. Losick.** 1995. Extracellular signal protein triggering the proteolytic activation of a developmental transcription factor in *B. subtilis*. Cell **83:**219-26.

45. **Jonas, R. M., E. A. Weaver, T. J. Kenney, C. P. Moran, Jr., and W. G. Haldenwang.** 1988. The *Bacillus subtilis spoIIG* operon encodes both σ^E and a gene necessary for σ^E activation. J. Bacteriol. **170:**507-11.

46. **Juan Wu, L., and J. Errington.** 2000. Identification and characterization of a new prespore-specific regulatory gene, *rsfA*, of *Bacillus subtilis*. J. Bacteriol. **182:**418-24.

47. **Karow, M. L., P. Glaser, and P. J. Piggot.** 1995. Identification of a gene, *spoIIR*, that links the activation of σ^E to the transcriptional activity of σ^F during sporulation in *Bacillus subtilis*. Proc. Natl. Acad. Sci. U. S. A. **92:**2012-6.

48. **Kay, D., and S. C. Warren.** 1968. Sporulation in *Bacillus subtilis*. Morphological changes. Biochem. J. **109:**819-24.

49. **Klaasen, H. L., J. P. Koopman, M. E. Van den Brink, M. H. Bakker, F. G. Poelma, and A. C. Beynen.** 1993. Intestinal, segmented, filamentous bacteria in a wide range of vertebrate species. Lab. Anim. **27:**141-50.

50. **Kroos, L., B. Kunkel, and R. Losick.** 1989. Switch protein alters specificity of RNA polymerase containing a compartment-specific sigma factor. Science **243:**526-9.

51. **LaBell, T. L., J. E. Trempy, and W. G. Haldenwang.** 1987. Sporulation-specific sigma factor σ^{29} of *Bacillus subtilis* is synthesized from a precursor protein, P^{31}. Proc. Natl. Acad. Sci. U. S. A. **84:**1784-8.

52. **Lagesen, K., P. Hallin, E. A. Rodland, H. H. Staerfeldt, T. Rognes, and D. W. Ussery.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35:**3100-8.

53. **Lechat, P., L. Hummel, S. Rousseau, and I. Moszer.** 2008. GenoList: an integrated environment for comparative analysis of microbial genomes. Nucleic Acids Res. **36:**D469-74.

54. **Liu, N. J., R. J. Dutton, and K. Pogliano.** 2006. Evidence that the SpoIIIE DNA translocase participates in membrane fusion during cytokinesis and engulfment. Mol. Microbiol. **59:**1097-113.

55. **Lopez, D., and R. Kolter.** 2010. Extracellular signals that define distinct and coexisting cell fates in *Bacillus subtilis*. FEMS Microbiol. Rev. **34:**134-49.

56. **Losick, R., and P. Stragier.** 1992. Crisscross regulation of cell-type-specific gene expression during development in *B. subtilis*. Nature **355:**601-4.

57. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25:**955-64.

58. **Marchler-Bauer, A., J. B. Anderson, M. K. Derbyshire, C. DeWeese-Scott, N. R. Gonzales, M. Gwadz, L. Hao, S. He, D. I. Hurwitz, J. D. Jackson, Z. Ke, D. Krylov, C. J. Lanczycki, C. A. Liebert, C. Liu, F. Lu, S. Lu, G. H. Marchler, M. Mullokandov, J. S. Song, N. Thanki, R. A. Yamashita, J. J. Yin, D. Zhang, and S. H. Bryant.** 2007. CDD: a conserved domain database for interactive domain family analysis. Nucleic Acids Res. **35:**D237-40.

59. **Margolis, P. S., A. Driks, and R. Losick.** 1993. Sporulation gene *spoIIB* from *Bacillus subtilis*. J. Bacteriol. **175:**528-40.

60. **McBride, S. M., A. Rubio, L. Wang, and W. G. Haldenwang.** 2005. Contributions of protein structure and gene position to the compartmentalization of the regulatory proteins σ^E and SpoIIE in sporulating *Bacillus subtilis*. Mol. Microbiol. **57:**434-51.

61. **Meisner, J., X. Wang, M. Serrano, A. O. Henriques, and C. P. Moran, Jr.** 2008. A channel connecting the mother cell and forespore during bacterial endospore formation. Proc. Natl. Acad. Sci. U. S. A. **105:**15100-5.

62. **Mendell, J. E., K. D. Clements, J. H. Choat, and E. R. Angert.** 2008. Extreme polyploidy in a large bacterium. Proc. Natl. Acad. Sci. U. S. A. **105:**6730-4.

63. **Miller, D. A., J. H. Choat, K. D. Clements, and E. R. Angert.** 2011. The *spoIIE* homolog of *Epulopiscium* sp. type B is expressed early in intracellular offspring development. J. Bacteriol. **193:**2642-6.

64. **Miller, D. A., G. Suen, D. Bruce, A. Copeland, J. F. Cheng, C. Detter, L. A. Goodwin, C. S. Han, L. J. Hauser, M. L. Land, A. Lapidus, S. Lucas, L. Meincke, S. Pitluck, R. Tapia, H. Teshima, T. Woyke, B. G. Fox, E. R. Angert, and C. R. Currie.** 2011. Complete genome sequence of the cellulose-degrading bacterium *Cellulosilyticum lentocellum*. J. Bacteriol. **193:**2357-8.

65. **Molle, V., M. Fujita, S. T. Jensen, P. Eichenberger, J. E. Gonzalez-Pastor, J. S. Liu, and R. Losick.** 2003. The Spo0A regulon of *Bacillus subtilis*. Mol. Microbiol. **50:**1683-701.

66. **Myers, E. W., G. G. Sutton, A. L. Delcher, I. M. Dew, D. P. Fasulo, M. J. Flanigan, S. A. Kravitz, C. M. Mobarry, K. H. Reinert, K. A. Remington, E. L. Anson, R. A. Bolanos, H. H. Chou, C. M. Jordan, A. L. Halpern, S. Lonardi, E. M. Beasley, R. C. Brandon, L. Chen, P. J. Dunn, Z. Lai, Y. Liang, D. R. Nusskern, M. Zhan, Q. Zhang, X. Zheng, G. M. Rubin, M. D. Adams, and J. C. Venter.** 2000. A whole-genome assembly of *Drosophila*. Science **287:**2196-204.

67. **Nicholson, W. L., N. Munakata, G. Horneck, H. J. Melosh, and P. Setlow.** 2000. Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments. Microbiol. Mol. Biol. Rev. **64:**548-72.

68. **Onyenwoke, R. U., J. A. Brill, K. Farahi, and J. Wiegel.** 2004. Sporulation genes in members of the low G+C Gram-type-positive phylogenetic branch (Firmicutes). Arch. Microbiol. **182:**182-92.

69. **Paredes, C. J., K. V. Alsaker, and E. T. Papoutsakis.** 2005. A comparative genomic view of clostridial sporulation and physiology. Nat. Rev. Microbiol. **3:**969-78.

70. **Partridge, S. R., and J. Errington.** 1993. The importance of morphological events and intercellular interactions in the regulation of prespore-specific gene expression during sporulation in *Bacillus subtilis*. Mol. Microbiol. **8:**945-55.

71. **Piggot, P. J., and J. G. Coote.** 1976. Genetic aspects of bacterial endospore formation. Bacteriol. Rev. **40:**908-62.

72. **Popham, D. L., and P. Stragier.** 1992. Binding of the *Bacillus subtilis spoIVCA* product to the recombination sites of the element interrupting the σ^K-encoding gene. Proc. Natl. Acad. Sci. U. S. A. **89:**5991-5.

73. **Ravagnani, A., K. C. Jennert, E. Steiner, R. Grunberg, J. R. Jefferies, S. R. Wilkinson, D. I. Young, E. C. Tidswell, D. P. Brown, P. Youngman, J. G. Morris, and M. Young.** 2000. Spo0A directly controls the switch from acid to solvent production in solvent-forming clostridia. Mol. Microbiol. **37:**1172-85.

74. **Ricca, E., S. Cutting, and R. Losick.** 1992. Characterization of *bofA*, a gene involved in intercompartmental regulation of pro-σ^K processing during sporulation in *Bacillus subtilis*. J. Bacteriol. **174:**3177-84.

75. **Robinow, C. F.** 1951. Observations on the structure of *Bacillus* spores. J. Gen. Microbiol. **5:**439-57.

76. **Ryter, A.** 1965. Etude morphologique de la sporulation de *Bacillus subtilis*. Ann. Inst. Pasteur. **108:**40-60.

77. **Saller, M. J., F. Fusetti, and A. J. Driessen.** 2009. *Bacillus subtilis* SpoIIIJ and YqjG function in membrane protein biogenesis. J. Bacteriol. **191:**6749-57.

78. **Setlow, B., K. A. McGinnis, K. Ragkousi, and P. Setlow.** 2000. Effects of major spore-specific DNA binding proteins on *Bacillus subtilis* sporulation and spore properties. J. Bacteriol. **182:**6906-12.

79. **Setlow, B., and P. Setlow.** 1995. Small, acid-soluble proteins bound to DNA protect *Bacillus subtilis* spores from killing by dry heat. Appl. Environ. Microbiol. **61:**2787-90.

80. **Setlow, P.** 2007. I will survive: DNA protection in bacterial spores. Trends Microbiol. **15:**172-80.

81. **Setlow, P.** 2006. Spores of *Bacillus subtilis*: their resistance to and killing by radiation, heat and chemicals. J. Appl. Microbiol. **101:**514-25.

82. **Sharp, M. D., and K. Pogliano.** 2003. The membrane domain of SpoIIIE is required for membrane fusion during *Bacillus subtilis* sporulation. J. Bacteriol. **185:**2005-8.

83. **Sharpe, M. E., and J. Errington.** 1995. Postseptational chromosome partitioning in bacteria. Proc. Natl. Acad. Sci. U. S. A. **92:**8630-4.

84. **Smith, L. D.** 1970. *Clostridium oceanicum*, sp. n., a sporeforming anaerobe isolated from marine sediments. J. Bacteriol. **103:**811-3.

85. **Steil, L., M. Serrano, A. O. Henriques, and U. Volker.** 2005. Genome-wide analysis of temporally regulated and compartment-specific gene expression in sporulating cells of *Bacillus subtilis*. Microbiology **151:**399-420.

86. **Steiner, E., A. E. Dago, D. I. Young, J. T. Heap, N. P. Minton, J. A. Hoch, and M. Young.** 2011. Multiple orphan histidine kinases interact directly with Spo0A to control the initiation of endospore formation in *Clostridium acetobutylicum*. Mol. Microbiol. **80:**641-54.

87. **Stephenson, K., and J. A. Hoch.** 2002. Evolution of signalling in the sporulation phosphorelay. Mol. Microbiol. **46:**297-304.

88. **Stragier, P.** 2002. A gene odyssey: exploring the genomes of endospore-forming bacteria, p. 519-525. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and its closest relatives. ASM Press, Washington D.C.

89. **Stragier, P., B. Kunkel, L. Kroos, and R. Losick.** 1989. Chromosomal rearrangement generating a composite gene for a developmental transcription factor. Science **243:**507-12.

90. **Sutton, G. G., O. White, M. D. Adams, and A. R. Kerlavage.** 1995. TIGR Assembler: A new tool for assembling large shotgun sequencing projects. Genome Sequencing Tech. **1:**9-19.

91. **Tatusov, R. L., M. Y. Galperin, D. A. Natale, and E. V. Koonin.** 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. **28:**33-6.

92. **van Ooij, C., P. Eichenberger, and R. Losick.** 2004. Dynamic patterns of subcellular protein localization during spore coat morphogenesis in *Bacillus subtilis*. J. Bacteriol. **186:**4441-8.

93. **Veening, J. W., L. W. Hamoen, and O. P. Kuipers.** 2005. Phosphatases modulate the bistable sporulation gene expression pattern in *Bacillus subtilis*. Mol. Microbiol. **56:**1481-94.

94. **Wang, S. T., B. Setlow, E. M. Conlon, J. L. Lyon, D. Imamura, T. Sato, P. Setlow, R. Losick, and P. Eichenberger.** 2006. The forespore line of gene expression in *Bacillus subtilis*. J. Mol. Biol. **358:**16-37.

95. **Ward, R. J., K. D. Clements, J. H. Choat, and E. R. Angert.** 2009. Cytology of terminally differentiated *Epulopiscium* mother cells. DNA Cell Biol. **28:**57-64.

96. **Wu, L. J., and J. Errington.** 1994. *Bacillus subtilis* SpoIIIE protein required for DNA segregation during asymmetric cell division. Science **264:**572-5.

97. **Zheng, L., R. Halberg, S. Roels, H. Ichikawa, L. Kroos, and R. Losick.** 1992. Sporulation regulatory protein GerE from *Bacillus subtilis* binds to and can activate or repress transcription from promoters for mother-cell-specific genes. J. Mol. Biol. **226:**1037-50.

CHAPTER 3

THE *SPOIIE* HOMOLOG OF *EPULOPISCIUM* SP. TYPE B IS EXPRESSED EARLY IN INTRACELLULAR OFFSPRING DEVELOPMENT*

* Adapted from Miller, D. A., J. H. Choat, K. D. Clements, and E. R. Angert. 2011. The *spoIIE* homolog of *Epulopiscium* sp. type B is expressed early in intracellular offspring development. Journal of Bacteriology. 193:2642-6. Copyright 2011. American Society for Microbiology

ABSTRACT

Epulopiscium sp. type B is an enormous intestinal symbiont of the surgeonfish *Naso tonganus*. Intracellular offspring production in *Epulopiscium* shares features with endospore formation. Here we characterize *spoIIE* in *Epulopiscium*. Timing of *spoIIE* expression and presence of interacting partners suggest that the activation of σ^F occurs early in *Epulopiscium* offspring development.

TEXT

Binary fission is an efficient means of reproduction, although alternative patterns of cell division and propagation have evolved in several bacterial lineages to better accommodate particular lifestyles (2). *Epulopiscium* spp. and their relatives comprise a morphologically diverse group of bacterial symbionts that inhabit the intestinal tract of surgeonfish (15, 23, 34). These Firmicutes exhibit an array of reproductive strategies that includes the formation of multiple intracellular offspring (15, 24). *Epulopiscium* sp. type B is the most extensively studied member of the group. These extraordinarily large, cigar-shaped cells can grow to lengths of 200-300 μm with widths of 50-60 μm (4). They are associated with the unicornfish *Naso tonganus* (Family Acanthuridae), and may aid in the breakdown of algae consumed by their host (15). *Epulopiscium* sp. type B reproduces solely by the formation of multiple internal offspring (Figure 3.1); generally two are produced but as many as five have been observed within a mother cell (15). In nature, intracellular offspring formation in *Epulopiscium* follows a recurrent daily cycle (34), leading to developmentally synchronized populations (4, 48).

Intracellular offspring formation in *Epulopiscium* spp. appears to have evolved from bacterial endospore formation (3), based on the evolutionary relationship of *Epulopiscium* with endospore-forming bacteria (5, 16) and morphologically defined stages that appear to be shared by the two processes (3-5). If intracellular offspring formation in *Epulopiscium* is related to endospore formation, we predict that genes involved early in sporulation will be conserved in *Epulopiscium* and used in offspring production (4).

Endospore development is best described in *Bacillus subtilis* (22, 44, 52). Progression through sporulation is driven in part by the sequential activation of alternative sigma factors. One key early, forespore-specific sigma factor is σ^F. Expression of the *spoIIA* operon

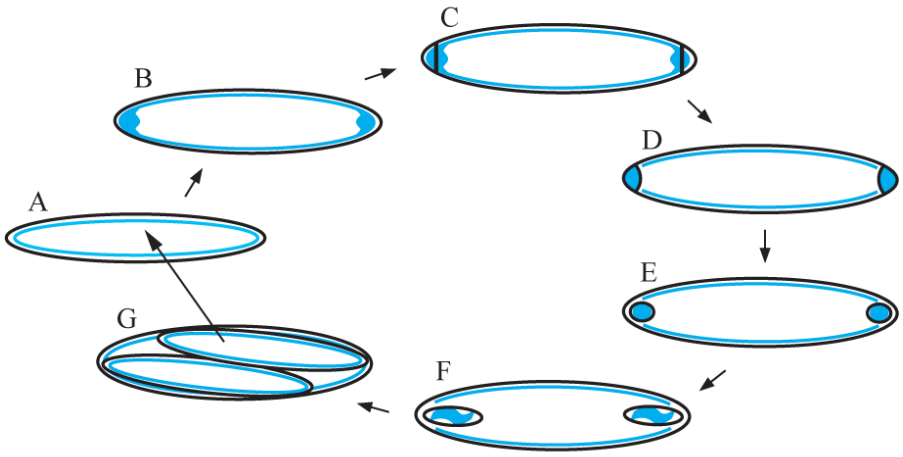


Figure 3.1. ***Epulopiscium*** **sp. type B life cycle.** Each major developmental transition is illustrated as a longitudinal section through a cell. (A) Prior to offspring initiation, DNA appears evenly distributed around the periphery of the cytoplasm. (B) As offspring formation begins, DNA accumulates at the poles of the cell. (C) The cell divides at both poles, trapping some of the polar DNA inside the newly formed offspring. (D) The remaining polar DNA is translocated inside the polar offspring. (E - G) Offspring are engulfed by the mother cell and elongate to fill the mother-cell cytoplasm. The offspring are released from the mother cell. In this drawing, DNA accumulation and polar division are shown in single cells, but in nature these transitions occur in offspring cells still contained within a mother cell. Cell outlines and division septa are shown in black and DNA in blue.

(comprised of *spoIIAA*, *spoIIAB* and *sigF*) occurs during the transition from exponential growth to stationary phase (49). The anti-sigma factor SpoIIAB holds σ^F inactive until polar division is complete (20, 33). Initially, the anti-anti sigma factor SpoIIAA is inactivated through phosphorylation by SpoIIAB (14, 19, 27, 33). SpoIIE dephosphorylates and thus activates SpoIIAA, which then disrupts the σ^F-SpoIIAB complex. Free SpoIIAB is rapidly degraded.

SpoIIE and FtsZ are binding partners that co-localize to the poles of the cell (8, 9, 31, 32, 36). Once asymmetric division is complete, SpoIIE accumulates in the polar septum (9). SpoIIE localization and its higher concentration within the smaller forespore direct forespore-specific activation of σ^F (6, 28-30).

The major vegetative growth sigma factor, σ^A, drives transcription of *spoIIE* (50), but prior to sporulation *spoIIE* expression is suppressed by Soj (37, 38). Transcription of *spoIIE* is upregulated by phosphorylated Spo0A and *spoIIE* transcripts can be detected approximately 1.5 hours after the initiation of sporulation (25, 50). All of these genes, *spoIIAA*, *spoIIAB*, *sigF*, *spoIIE*, *spo0A* and *soj*, are conserved in endospore-forming bacteria and essential for sporulation (18, 37, 41, 44).

Structural analysis of the *Epulopiscium* SpoIIE

Putative homologs of *spoIIE*, *spoIIAA*, *spoIIAB* and *sigF* have been identified in the *Epulopiscium* sp. type B genome, which suggests *Epulopiscium* has an early offspring-specific sigma factor activated by the cascade of molecular interactions described for *B. subtilis*. SpoIIE is a large protein with three distinct domains. The N-terminus targets SpoIIE to the cell membrane (7, 9). Domain II facilitates interaction with FtsZ and SpoIIE oligomerization (32). Conserved amino acid residues in the C-terminal phosphatase domain place SpoIIE within the PP2C protein phosphatase family (1, 42). All strains, plasmids and primers used in the following study are provided in Tables 3.1 & 3.2. The sequence of *spoIIE* from *Epulopiscium* plus 269 bases upstream of the predicted start codon was determined (GenBank accession number HQ149097). The gene is 2,352 bp long and the predicted protein shares 22% amino acid identity and 50% similarity with *B. subtilis* SpoIIE (Figure 3.2). A hydrophobicity plot of *Epulopiscium* SpoIIE reveals a potential N-terminus transmembrane domain to approximately residue 304,

Table 3.1. Strains and plasmids used in this study

Strain or Plasmid	Relevant Characteristics or Purpose	Source or Reference
Strains		
E. coli DH5α	Plasmid cloning	Laboratory stock
E. coli TOP10	Plasmid cloning	Invitrogen
B. subtilis PY79	Lab strain, sporulation	(51)
Plasmids		
pRSET IIE-GFP	*Epulopiscium spoIIE-gfp* in pRSETB	This study
pSWEET2E	*B. subtilis spoIIE* in pSWEET	This study
pDM1	*Epulopiscium rpoB* fragment in pCR 2.1	This study
pDM2	*B. subtilis rpoB* fragment in pCR 2.1	This study
pRSETB	Amp^r, cloning vector	Invitrogen
pSWEET	Amp^r, Cm^r, cloning vector	(10)
pCR 2.1-TOPO	Amp^r, Kan^r, cloning vector	Invitrogen

Table 3.2. Primers used in this study

Primer	Designation	Sequence (5´ - 3´)
1	GFPtagF	GGGAGTGGAAGCTTGGATGAGTAAAGGAGAAGAACTTTTC
2	GFPtagR	GGGAGTGGAAGCTTTTTGTATAGTTCATCCATGCCATG
3	SpoIIEABfragF	AGTACCAAACGCGAAGAACAAAAT
4	SpoIIEABfragR	ATAGTCTCCAGCAAGAAGGCTTCGAC
5	BsubspoIIEcompF	GCGGCGTTAATTAAGGGAAAAGGTGGTGAACTACTATGGAAAAAGCAGAAAGAAGAG
6	BsubspoIIEcompR	CCTGCGCTAGCTTATGAAATTTCTTGTTTGTTTTGAAAGAT
7	EpulorpoB742F	CCACCAACGGTTGAAAGTGCTGAA
8	EpulorpoB1107R	CGATGATTTCTGCCAACACTGG
9	BsubrpoB3503R	CCGTCAGCTTGTTTCGCATCTTCT
10	BsubrpoB3064F	ACTGGAGAGCCGTTTGATAACCGT
11	EpulorpoB913F	GAAGCTGGTGATAAAATTTCTGAAG
12	EpulorpoB994R	CAATCTTTACATTTGAAGTCCCAGTA
13	BsubrpoB3183F	GCAGCCTCTTGGCGGTAA
14	BsubrpoB3243R	CCAAACCTCCATCTCACCAAA
15	EpulospoIIE1203F	GGTTTATAAAGGAGAACGCAATGG
16	EpulospoIIE1280R	GCAGAGGGACATTTATTACTGAATGA
17	BsubspoIIE302F	TGCTCATACTGGCGGCATT
18	BsubspoIIE358R	TGAAGGCAGCCACTTTAGAAAAT

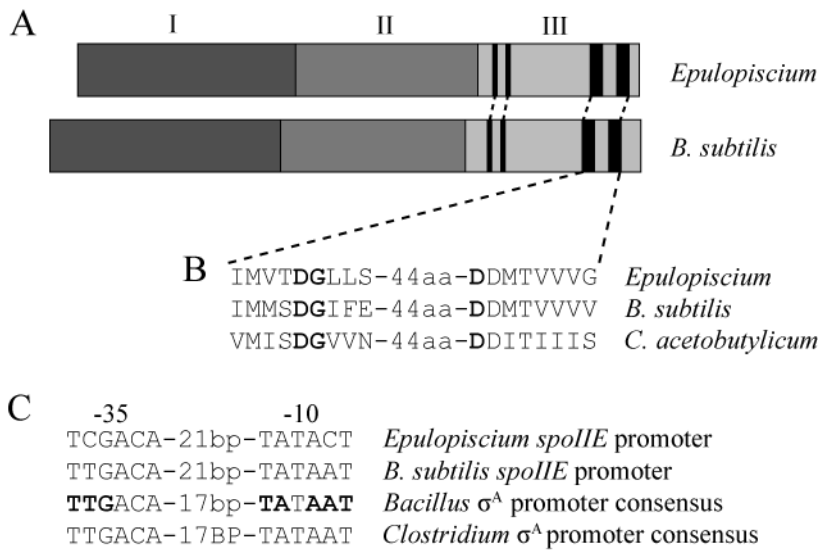


Figure 3.2. Analysis of the putative *spoIIE* homolog from *Epulopiscium* sp. type B. (A) Domain structure of SpoIIE from *Epulopiscium* and *B. subtilis*. Black bars in Domain III indicate active site residues (thin) and motifs (thick) found in PP2C phosphatases. (B) Conserved PP2C motifs from *Epulopiscium*, *B. subtilis* and *Clostridium acetobutylicum*. Bold font indicates residues conserved in all PP2C protein phosphatases. (C) Comparison of *Epulopiscium* and *B. subtilis spoIIE* promoters with σ^A consensus sequences found in *B. subtilis* and *Clostridium* spp. Bold font indicates conserved bases in the *B. subtilis* consensus (26).

similar to the 324-amino acid hydrophobic domain I of *B. subtilis* SpoIIE (data not shown) (7, 9). Domain II is not highly conserved between known SpoIIE homologs, and the *B. subtilis* and *Epulopiscium* proteins are only 16% identical. The phosphatase domain (III) is highly conserved, with 35% identity and 64% similarity to *B. subtilis* SpoIIE. Two essential active site aspartic acid residues and two downstream amino acid motifs (Figure 3.2B) found in all PP2C type phosphatases (1) are conserved in *Epulopiscium* SpoIIE. The σ^A promoters of both *B. subtilis* and *Epulopiscium spoIIE* have an unusual spacing of 21 bp (instead of the typical 17 bp)

between the -10 and -35 sequences (50).

Genes for *spoIIAA*, *spoIIAB* and *sigF*, are found in an operon in the *Epulopiscium* genome. Alignments of the predicted protein sequences of SpoIIAA, SpoIIAB, and SigF from *Epulopiscium* with *B. subtilis* homologs indicate 27%, 41% and 48% identity and 66%, 67% and 72% similarity, respectively.

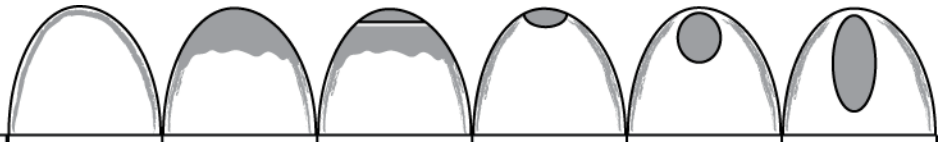
Developmentally synchronized populations of *Epulopiscium*.

Naso tonganus, which feeds diurnally on plant material but retains food in the alimentary tract overnight (11), were collected by spearfishing on outer reefs in the vicinity of Lizard Island, Great Barrier Reef, Australia between 1000-1400 hrs. Ethanol-fixed intestinal contents were stained with 2 µg/ml DAPI (48), and randomly selected *Epulopiscium* cells were classified by stage of development (Figure 3.3). Since morphological transitions in *Epulopiscium* are similar to those in a sporulating cell (4), we used the classically described stages of endospore formation (40) to represent similar events in *Epulopiscium* offspring development.

Expression of *spoIIE* peaks early in offspring development.

Intestinal contents from *N. tonganus* were also fixed with RNAProtect (Qiagen). For each sample, total RNA isolated from a 1 ml aliquot was converted to cDNA and used in quantitative PCR assays using Power SyBER Green Master Mix (Applied Biosystems). All qPCR primers were designed using Primer Express Software and qPCR was performed using the ABI 7300 Real Time PCR System with default cycling conditions. To control for differences in *Epulopiscium* cell density in these samples, the housekeeping gene *rpoB* was used to normalize the data. This approach has been used in other studies of population-specific gene expression (12, 17, 39, 45) and seemed appropriate here; based on previously published microarray studies *rpoB* expression appears constant throughout sporulation in *B. subtilis* and is not affected by σ^H,

Figure 3.3. Characterization of *Epulopiscium* cell development in *N. tonganus* gut samples. Drawings of the developmental stages are shown across the top. Sample number and the total number of categorized cells are show on the left. The number of cells and percentage of the sample representing any given stage are provided. Stages are defined as follows. Stage 0: a cell that contains large offspring that show no signs of next-round offspring initiation. Stage I cells have large amounts of coalesced polar DNA but no septa. Stage II cells have polar septa but not all polar DNA is inside the polar cells. Stage II - III cells have DNA translocated inside the polar cell; the polar septa are curved, indicating engulfment. Stage III and stage IV* contain cells with engulfed offspring, but stage III cells have an offspring length-to-width ratio of less than 2:1 while stage IV* cells have a greater ratio. Some of the hallmarks of stage IV forespores, such as the formation of cortex and assembly of coat, may not occur in *Epulopiscium* cells. To highlight these differences we refer to this stage as IV* in *Epulopiscium*.



Sample Total Tallied	**0**	**I**	**II**	**II-III**	**III**	**IV***
1 325 cells	0	71.1% 231	25.8% 84	2.8% 9	0.3% 1	0
2 327 cells	0.3% 1	60.6% 198	20.8% 68	17.1% 56	1.2% 4	0
3 449 cells	0	2.2% 10	11.6% 52	32.3% 145	53.9% 242	0
4 318 cells	0	0	0	0.6% 2	63.2% 201	36.3% 115
5 339 cells	0.9% 3	0	0	0	0	99.1% 336
6 352 cells	0.6% 2	0	0	0	0	99.4% 350

Spo0A, σ^F, σ^E, σ^G or σ^K (21, 47). Specificity of all qPCR products generated from intestinal samples was confirmed by sequence analysis. Samples 1 and 2 had the highest *spoIIE*-to-*rpoB* transcript ratios (0.2256 and 0.1452), indicating that *spoIIE* is highly expressed prior to polar division (Figure 3.4A). Sample 3 had a lower transcript ratio (0.0560), which correlates well with the distribution of the cells that had advanced beyond polar septum formation. Samples 4, 5 and 6 had low transcript ratios (0.0021, 0.0035 and 0.0020, respectively), which is consistent with these post-engulfment populations.

To the best of our knowledge, an RT-qPCR assay for *spoIIE* expression has never been reported for sporulating *B. subtilis*, and thus we wanted to validate this method. Samples of *B. subtilis* were taken every half hour from the onset of sporulation (t_0), induced by resuspension (35), and assayed (Figure 3.4B). No-template controls for *spoIIE* yielded values that were three orders of magnitude less than any sample. Prior to t_1 a low *spoIIE*-to-*rpoB* transcript ratio was observed (0.0067 for t_0, 0.0052 for $t_{0.5}$) transcript ratios then increased for t_1 and $t_{1.5}$ (0.0219, 0.0214, respectively). Samples from t_2 and later had greater ratios (0.2466, 0.1828, 0.3143, and 0.1425). Microscopic examination of cells stained with FM 4-64, Mitotracker green and DAPI (43) was used to determine the proportion of cells at specific morphological stages of sporulation (no polar septum, straight polar septum, curved septum, completely engulfed forespore) at each time point (data not shown). Large numbers of cells with a polar septum begin to appear at $t_{1.5}$ (8.2% of the population) and this stage becomes more prominent at t_2 (25.6%). By t_3, cells with a fully engulfed forespore are more abundant than cells with a polar septum. The qPCR results are comparable to published *spoIIE* transcription profiles using β-galactosidase fusions (25), where *spoIIE* expression increases considerably in populations with cells that are just beginning

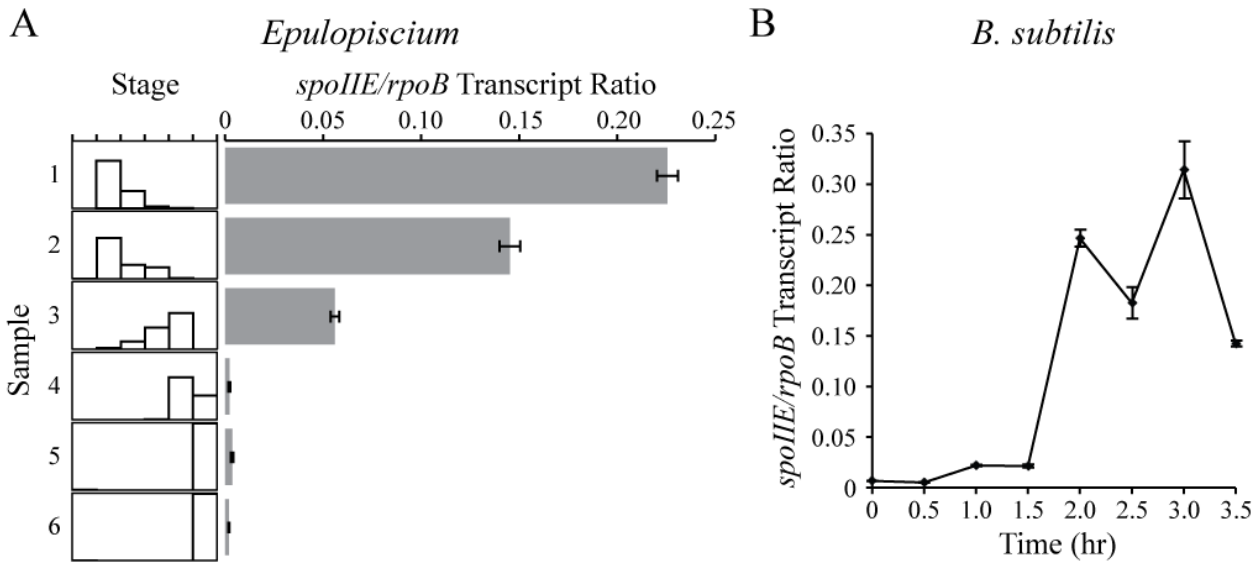


Figure 3.4. *spoIIE* expression in *Epulopiscium* and sporulating *B. subtilis*. (A) For *Epulopiscium*, the bar graph on the left represents stages of development (from Figure 3.3). The grey horizontal bars indicate the ratio of *spoIIE* transcripts to *rpoB* transcripts. (B) For *B. subtilis*, *spoIIE*-to-*rpoB* transcript ratios were determined for subsamples taken every 30 min throughout the first 3.5 hours of sporulation. Error bars represent the standard error of the means of the four unique *spoIIE*-to-*rpoB* pairings of the data generated from duplicate RT-qPCR reactions.

to divide asymmetrically and persists even when the population has progressed and engulfed cells are abundant. When compared with *Epulopiscium*, the timing of *spoIIE* expression appears slightly later in developmental progression in *B. subtilis* populations and expression persists for a longer developmental interval. This may reflect response heterogeneity of sporulating *B. subtilis* cultures (13, 46). Sporulation is a last resort and there is a clear advantage to the population for some cells to delay the commitment to sporulate as long as possible. Conversely, intracellular offspring formation is essential for *Epulopiscium* reproduction and synchronized development may presage recurrent fluctuations in nutrients (48), thus there would be little advantage for some cells in the population to delay development.

***spoIIE* expression patterns suggest a role in early offspring development**

Taken together, the RT-qPCR results from both *Epulopiscium* and *B. subtilis* indicate that the *spoIIE* homolog in *Epulopiscium* is expressed in a manner that would support its predicted role in offspring development. Expression peaks just prior to polar division, when SpoIIE would be needed for polar septum formation and subsequent activation of σ^F. The use of cell-specific sigma factors could drive the alternative cell fates of the offspring and the mother cell in *Epulopiscium*, allowing for the growth of the offspring while the mother cell initially supports offspring growth but eventually progresses toward apoptosis (48). Further investigation of these potential transcription programs may indicate where the line is drawn between intracellular offspring formation and sporulation.

REFERENCES

1. **Adler, E., A. Donella-Deana, F. Arigoni, L. A. Pinna, and P. Stragier.** 1997. Structural relationship between a bacterial developmental protein and eukaryotic PP2C protein phosphatases. Mol. Microbiol. **23:**57-62.
2. **Angert, E. R.** 2005. Alternatives to binary fission in bacteria. Nat. Rev. Microbiol. **3:**214-24.
3. **Angert, E. R., A. E. Brooks, and N. R. Pace.** 1996. Phylogenetic analysis of *Metabacterium polyspora*: clues to the evolutionary origin of daughter cell production in *Epulopiscium* species, the largest bacteria. J. Bacteriol. **178:**1451-6.
4. **Angert, E. R., and K. D. Clements.** 2004. Initiation of intracellular offspring in *Epulopiscium*. Mol. Microbiol. **51:**827-35.
5. **Angert, E. R., K. D. Clements, and N. R. Pace.** 1993. The largest bacterium. Nature **362:**239-41.
6. **Arigoni, F., L. Duncan, S. Alper, R. Losick, and P. Stragier.** 1996. SpoIIE governs the phosphorylation state of a protein regulating transcription factor σ^F during sporulation in *Bacillus subtilis*. Proc. Natl. Acad. Sci. U. S. A. **93:**3238-42.
7. **Arigoni, F., A. M. Guerout-Fleury, I. Barak, and P. Stragier.** 1999. The SpoIIE phosphatase, the sporulation septum and the establishment of forespore-specific transcription in *Bacillus subtilis*: a reassessment. Mol. Microbiol. **31:**1407-15.
8. **Arigoni, F., K. Pogliano, C. D. Webb, P. Stragier, and R. Losick.** 1995. Localization of protein implicated in establishment of cell type to sites of asymmetric division. Science **270:**637-40.

9. **Barak, I., J. Behari, G. Olmedo, P. Guzman, D. P. Brown, E. Castro, D. Walker, J. Westpheling, and P. Youngman.** 1996. Structure and function of the *Bacillus* SpoIIE protein and its localization to sites of sporulation septum assembly. Mol. Microbiol. **19:**1047-60.

10. **Bhavsar, A. P., X. Zhao, and E. D. Brown.** 2001. Development and characterization of a xylose-dependent system for expression of cloned genes in *Bacillus subtilis*: conditional complementation of a teichoic acid mutant. Appl. Environ. Microbiol. **67:**403-10.

11. **Choat, J. H., W. D. Robbins, and K. D. Clements.** 2004. The trophic status of herbivorous fishes on coral reefs. II. Food processing modes and trophodynamics. Marine Biol. **145:**445-454.

12. **Chow, W. L., D. Cheng, S. Wang, and J. He.** 2010. Identification and transcriptional analysis of *trans*-DCE-producing reductive dehalogenases in *Dehalococcoides* species. ISME J. **4:**1020-30.

13. **Chung, J. D., G. Stephanopoulos, K. Ireton, and A. D. Grossman.** 1994. Gene expression in single cells of *Bacillus subtilis*: evidence that a threshold mechanism controls the initiation of sporulation. J. Bacteriol. **176:**1977-84.

14. **Clarkson, J., I. D. Campbell, and M. D. Yudkin.** 2004. Physical evidence for the induced release of the *Bacillus subtilis* transcription factor, σ^F, from its inhibitory complex. J. Mol. Biol. **340:**203-9.

15. **Clements, K. D., D. C. Sutton, and J. H. Choat.** 1989. Occurrence and characteristics of unusual protistan symbionts from surgeonfishes Acanthuridae of the Great Barrier Reef Australia. Marine Biol. **102:**403-12.

16. **Collins, M. D., P. A. Lawson, A. Willems, J. J. Cordoba, J. Fernandez-Garayzabal, P. Garcia, J. Cai, H. Hippe, and J. A. Farrow.** 1994. The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. Int. J. Syst. Bacteriol. **44:**812-26.

17. **Doumith, M., M. J. Ellington, D. M. Livermore, and N. Woodford.** 2009. Molecular mechanisms disrupting porin expression in ertapenem-resistant *Klebsiella* and *Enterobacter* spp. clinical isolates from the UK. J. Antimicrob. Chemother. **63:**659-67.

18. **Duncan, L., S. Alper, F. Arigoni, R. Losick, and P. Stragier.** 1995. Activation of cell-specific transcription by a serine phosphatase at the site of asymmetric division. Science **270:**641-4.

19. **Duncan, L., S. Alper, and R. Losick.** 1996. SpoIIAA governs the release of the cell-type specific transcription factor σ^F from its anti-sigma factor SpoIIAB. J. Mol. Biol. **260:**147-64.

20. **Duncan, L., and R. Losick.** 1993. SpoIIAB is an anti-σ factor that binds to and inhibits transcription by regulatory protein σ^F from *Bacillus subtilis*. Proc. Natl. Acad. Sci. U. S. A. **90:**2325-9.

21. **Eichenberger, P., M. Fujita, S. T. Jensen, E. M. Conlon, D. Z. Rudner, S. T. Wang, C. Ferguson, K. Haga, T. Sato, J. S. Liu, and R. Losick.** 2004. The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. PLoS Biol. **2:**e328.

22. **Errington, J.** 2003. Regulation of endospore formation in *Bacillus subtilis*. Nat. Rev. Microbiol. **1:**117-26.

23. **Fishelson, L., W. L. Montgomery, and A. A. Myrberg.** 1985. A Unique Symbiosis in the Gut of Tropical Herbivorous Surgeonfish (Acanthuridae: Teleostei) from the Red Sea. Science **229:**49-51.

24. **Flint, J. F., D. Drzymalski, W. L. Montgomery, G. Southam, and E. R. Angert.** 2005. Nocturnal production of endospores in natural populations of *Epulopiscium*-like surgeonfish symbionts. J. Bacteriol. **187:**7460-70.

25. **Guzman, P., J. Westpheling, and P. Youngman.** 1988. Characterization of the promoter region of the *Bacillus subtilis spoIIE* operon. J. Bacteriol. **170:**1598-609.

26. **Helmann, J. D.** 1995. Compilation and analysis of Bacillus subtilis σ^A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. Nucleic Acids Res. **23:**2351-60.

27. **Ho, M. S., K. Carniol, and R. Losick.** 2003. Evidence in support of a docking model for the release of the transcription factor σ^F from the antisigma factor SpoIIAB in *Bacillus subtilis*. J. Biol. Chem. **278:**20898-905.

28. **Iber, D.** 2006. A computational analysis of the impact of the transient genetic imbalance on compartmentalized gene expression during sporulation in *Bacillus subtilis*. J. Mol. Biol. **360:**15-20.

29. **Iber, D., J. Clarkson, M. D. Yudkin, and I. D. Campbell.** 2006. The mechanism of cell differentiation in *Bacillus subtilis*. Nature **441:**371-4.

30. **King, N., O. Dreesen, P. Stragier, K. Pogliano, and R. Losick.** 1999. Septation, dephosphorylation, and the activation of σ^F during sporulation in *Bacillus subtilis*. Genes Dev. **13:**1156-67.

31. **Levin, P. A., R. Losick, P. Stragier, and F. Arigoni.** 1997. Localization of the sporulation protein SpoIIE in *Bacillus subtilis* is dependent upon the cell division protein FtsZ. Mol. Microbiol. **25:**839-46.

32. **Lucet, I., A. Feucht, M. D. Yudkin, and J. Errington.** 2000. Direct interaction between the cell division protein FtsZ and the cell differentiation protein SpoIIE. EMBO J. **19:**1467-75.

33. **Min, K. T., C. M. Hilditch, B. Diederich, J. Errington, and M. D. Yudkin.** 1993. σ^F, the first compartment-specific transcription factor of *B. subtilis*, is regulated by an anti-σ factor that is also a protein kinase. Cell **74:**735-42.

34. **Montgomery, W. L., and P. E. Pollak.** 1988. *Epulopiscium fishelsoni* n.g., n.s., a protist of uncertain taxonomic affinities from the gut of an herbivorous reef fish. J. Protozool. **35:**565-69.

35. **Nicholson, W. L., and P. Setlow.** 1990. Sporulation, Germination and Outgrowth, p. 396-399, 432. *In* C. R. Harwood and S. M. Cutting (ed.), Molecular Biological Methods for *Bacillus*. John Whiley and Sons Ltd, Chichester, England.

36. **Prepiak, P., Z. Chromikova, and I. Barak.** 2001. Use of yeast two-hybrid system for detection of *Bacillus subtilis* FtsZ protein partners. Folia Microbiol. **46:**292-6.

37. **Quisel, J. D., and A. D. Grossman.** 2000. Control of sporulation gene expression in *Bacillus subtilis* by the chromosome partitioning proteins Soj (ParA) and Spo0J (ParB). J. Bacteriol. **182:**3446-51.

38. **Quisel, J. D., D. C. Lin, and A. D. Grossman.** 1999. Control of development by altered localization of a transcription factor in *B. subtilis*. Mol. Cell **4:**665-72.

39. **Rahm, B. G., R. M. Morris, and R. E. Richardson.** 2006. Temporal expression of respiratory genes in an enrichment culture containing *Dehalococcoides ethenogenes*. Appl. Environ. Microbiol. **72:**5486-91.

40. **Ryter, A.** 1965. Etude morphologique de la sporulation de *Bacillus subtilis*. Ann. Inst. Pasteur. **108:**40-60.

41. **Schmidt, R., P. Margolis, L. Duncan, R. Coppolecchia, C. P. Moran, Jr., and R. Losick.** 1990. Control of developmental transcription factor σ^F by sporulation regulatory proteins SpoIIAA and SpoIIAB in *Bacillus subtilis*. Proc. Natl. Acad. Sci. U. S. A. **87:**9221-5.

42. **Schroeter, R., S. Schlisio, I. Lucet, M. Yudkin, and R. Borriss.** 1999. The *Bacillus subtilis* regulator protein SpoIIE shares functional and structural similarities with eukaryotic protein phosphatases 2C. FEMS Microbiol. Lett. **174:**117-23.

43. **Sharp, M. D., and K. Pogliano.** 1999. An *in vivo* membrane fusion assay implicates SpoIIIE in the final stages of engulfment during *Bacillus subtilis* sporulation. Proc. Natl. Acad. Sci. U. S. A. **96:**14553-8.

44. **Stragier, P., and R. Losick.** 1996. Molecular genetics of sporulation in *Bacillus subtilis*. Annu. Rev. Genet. **30:**297-41.

45. **Szabo, D., F. Silveira, A. M. Hujer, R. A. Bonomo, K. M. Hujer, J. W. Marsh, C. R. Bethel, Y. Doi, K. Deeley, and D. L. Paterson.** 2006. Outer membrane protein changes and efflux pump expression together may confer resistance to ertapenem in *Enterobacter cloacae*. Antimicrob. Agents Chemother. **50:**2833-5.

46. **Veening, J. W., L. W. Hamoen, and O. P. Kuipers.** 2005. Phosphatases modulate the bistable sporulation gene expression pattern in *Bacillus subtilis*. Mol. Microbiol. **56:**1481-94.

47. **Wang, S. T., B. Setlow, E. M. Conlon, J. L. Lyon, D. Imamura, T. Sato, P. Setlow, R. Losick, and P. Eichenberger.** 2006. The forespore line of gene expression in *Bacillus subtilis*. J. Mol. Biol. **358:**16-37.

48. **Ward, R. J., K. D. Clements, J. H. Choat, and E. R. Angert.** 2009. Cytology of terminally differentiated *Epulopiscium* mother cells. DNA Cell Biol. **28:**57-64.

49. **Wu, J. J., P. J. Piggot, K. M. Tatti, and C. P. Moran, Jr.** 1991. Transcription of the *Bacillus subtilis spoIIA* locus. Gene **101:**113-6.

50. **York, K., T. J. Kenney, S. Satola, C. P. Moran, Jr., H. Poth, and P. Youngman.** 1992. Spo0A controls the σ^A-dependent activation of *Bacillus subtilis* sporulation-specific transcription unit *spoIIE*. J. Bacteriol. **174:**2648-58.

51. **Youngman, P., J. B. Perkins, and R. Losick.** 1984. Construction of a cloning site near one end of Tn917 into which foreign DNA may be inserted without affecting transposition in *Bacillus subtilis* or expression of the transposon-borne *erm* gene. Plasmid **12:**1-9.

52. **Yudkin, M. D., and J. Clarkson.** 2005. Differential gene expression in genetically identical sister cells: the initiation of sporulation in *Bacillus subtilis*. Mol. Microbiol. **56:**578-89.

APPENDIX 1

COMPLETE GENOME SEQUENCE OF THE CELLULOSE-DEGRADING BACTERIUM *CELLULOSILYTICUM LENTOCELLUM**

* Adapted from Miller, D. A., G. Suen, D. Bruce, A. Copeland, J. F. Cheng, C. Detter, L. A. Goodwin, C. S. Han, L. J. Hauser, M. L. Land, A. Lapidus, S. Lucas, L. Meincke, S. Pitluck, R. Tapia, H. Teshima, T. Woyke, B. G. Fox, E. R. Angert, and C. R. Currie. 2011. Complete genome sequence of the cellulose-degrading bacterium *Cellulosilyticum lentocellum*. Journal of Bacteriology. 193:2357-8. Copyright 2011. American Society for Microbiolgy.

ABSTRACT

Cellulosilyticum lentocellum DSM 5427 is an anaerobic, endospore-forming member of the Firmicutes. We describe the complete genome sequence of this cellulose-degrading bacterium; originally isolated from estuarine sediment of a river that received both domestic and paper mill waste. Comparative genomics of cellulolytic clostridia will provide insight into factors that influence degradation rates.

TEXT

Cellulosilyticum lentocellum DSM 5427 (2), previously known as *Clostridium lentocellum*, was isolated from estuarine sediment at the mouth of the River Don, Aberdeenshire, Scotland (15). The sample was sourced for potential novel cellulose degraders because the river received both domestic and paper-mill effluent and its sediments exhibited cellulolytic activity. *Cellulosilyticum lentocellum* is able to degrade cellulose slowly and may form a single terminal endospore (15). Based on 16S rRNA gene sequence comparisons, *C. lentocellum* DSM 5427 belongs to clostridium cluster XIVb (4), or *Lachnospiraceae*, with its closest relatives being *Cellulosilyticum rumincola* (2), *Metabacterium polyspora* and *Epulopiscium* spp. (2, 4). The phylogenetic relationship between *C. lentocellum* DSM 5427 and *Clostridium lentocellum* SG6 (16, 17) remains unclear, as no gene sequence data are available from the latter isolate and it has not been deposited in any public culture collection. *Clostridium lentocellum* SG6 was isolated from budgerigar bird droppings and named based on its phenotypic similarities to *C. lentocellum* DSM 5427.

The *C. lentocellum* DSM 5427 genome sequence was determined by the Joint Genome Institute (JGI) using a combination of 454 Titanium (14) and Illumina (1) technologies. General descriptions of library construction, sequencing, and assembly can be found on the JGI website at http://www.jgi.doe.gov/. Illumina sequencing data was assembled with VELVET (19), and the consensus sequences were shredded into 1.5 kb overlapped fake reads and assembled together with the 454 data. The initial Newbler assembly contained 125 contigs in 12 scaffolds. The initial 454 assembly was converted into a phrap assembly by making fake reads from the consensus and collecting the read pairs in the 454 paired end library. The Phred/Phrap/Consed software package (5-7) was used for sequence assembly and quality assessment as follows.

Illumina data was used to correct potential base errors and increase consensus quality using the software Polisher developed at the JGI (Alla Lapidus, unpublished). After the shotgun stage, reads were assembled with parallel Phrap (High Performance Software, LLC). Possible mis-assemblies were corrected with gapResolution (Cliff Han, unpublished), Dupfinisher (8), or sequencing cloned bridging PCR fragments with subcloning. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR primer walks. A total of 364 additional reactions and were necessary to close gaps and to raise the quality of the finished sequence. Annotation of this genome was performed at Oak Ridge National Laboratory (ORNL) using an automated pipeline that includes open reading frame prediction using PRODIGAL (10), and functional annotation using Clusters of Orthologous Genes (COG) (18) and KEGG (11). Genes encoding tRNAs and rRNA operons were determined using tRNAscan-SE (13) and RNAmmer 1.2 (12), respectively.

The genome of *C. lentocellum* DSM 5427 consists of a single circular chromosome of 4,714,237 bp with a G+C content of 34.3%. The genome contains 4,185 predicted protein-coding sequences, 105 tRNAs, twelve 23S rRNAs and eleven 16S rRNAs. COG annotation showed 9.74% of predicted protein coding sequences fall into the carbohydrate transport and metabolism category. Thirteen predicted ORFs are annotated as cellulose degradation enzymes. Seven of these coding sequences are most closely related to cellulases in *C. rumincola* (3). This is the first completely assembled genome sequence from a member of the clostridium cluster XIVb, a diverse group of anaerobes isolated from anoxic sediments or gastrointestinal samples from herbivores. Comparative studies with other cellulose degraders (9) holds promise for advancing our understanding of the cellular and genomic factors that influence cellulose degradation and biofuel production.

Nucleotide sequence accession number. The genome sequence of *C. lentocellum* DSM 5427 has been deposited in GenBank under accession number ADVF00000000.

REFERENCES

1. **Bennett, S.** 2004. Solexa Ltd. Pharmacogenomics **5:**433-8.
2. **Cai, S., and X. Dong.** 2010. *Cellulosilyticum ruminicola* gen. nov., sp. nov., isolated from the rumen of yak, and reclassification of *Clostridium lentocellum* as *Cellulosilyticum lentocellum* comb. nov. Int. J. Syst. Evol. Microbiol. **60:**845-9.
3. **Cai, S., J. Li, F. Z. Hu, K. Zhang, Y. Luo, B. Janto, R. Boissy, G. Ehrlich, and X. Dong.** 2010. *Cellulosilyticum ruminicola*, a newly described rumen bacterium that possesses redundant fibrolytic-protein-encoding genes and degrades lignocellulose with multiple carbohydrate- borne fibrolytic enzymes. Appl. Environ. Microbiol. **76:**3818-24.
4. **Collins, M. D., P. A. Lawson, A. Willems, J. J. Cordoba, J. Fernandez-Garayzabal, P. Garcia, J. Cai, H. Hippe, and J. A. Farrow.** 1994. The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. Int. J. Syst. Bacteriol. **44:**812-26.
5. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:**186-94.
6. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8:**175-85.
7. **Gordon, D., C. Abajian, and P. Green.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8:**195-202.
8. **Han, C. S., and P. Chain.** 2006. Finishing repeat regions automatically with Dupfinisher, p. 141-146. *In* H. R. Arabnia. and H. Valafar (ed.), Proceedings of the 2006 International Conference on Bioinformatics and Computational Biology. CSREA Press, Las Vegas, NV.

9. **Hemme, C. L., H. Mouttaki, Y. J. Lee, G. Zhang, L. Goodwin, S. Lucas, A. Copeland, A. Lapidus, T. Glavina del Rio, H. Tice, E. Saunders, T. Brettin, J. C. Detter, C. S. Han, S. Pitluck, M. L. Land, L. J. Hauser, N. Kyrpides, N. Mikhailova, Z. He, L. Wu, J. D. Van Nostrand, B. Henrissat, Q. He, P. A. Lawson, R. S. Tanner, L. R. Lynd, J. Wiegel, M. W. Fields, A. P. Arkin, C. W. Schadt, B. S. Stevenson, M. J. McInerney, Y. Yang, H. Dong, D. Xing, N. Ren, A. Wang, R. L. Huhnke, J. R. Mielenz, S. Y. Ding, M. E. Himmel, S. Taghavi, D. van der Lelie, E. M. Rubin, and J. Zhou.** Sequencing of multiple clostridial genomes related to biomass conversion and biofuel production. J. Bacteriol. **192:**6494-6.

10. **Hyatt, D., G. L. Chen, P. F. Locascio, M. L. Land, F. W. Larimer, and L. J. Hauser.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119.

11. **Kanehisa, M., S. Goto, S. Kawashima, Y. Okuno, and M. Hattori.** 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res. **32:**D277-80.

12. **Lagesen, K., P. Hallin, E. A. Rodland, H. H. Staerfeldt, T. Rognes, and D. W. Ussery.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35:**3100-8.

13. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25:**955-64.

14. **Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:**376-80.

15. **Murray, W. D., L. Hoffman, N. L. Campbell, and R. H. Madden.** 1986. *Clostridium lentocellum* sp. nov., a cellulolytic species from river sediment conatining paper-mill waste. Syst. Appl. Microbiol. **8:**181-184.

16. **Ravinder, T., K. Sudha Rani, R. Gopal, and G. Seenayya.** 1998. Direct conversion of biomass to acetic acid by anaerobic cellulolytic isolates. J. Sci. Ind. Res. **57:**591-594.

17. **Ravinder, T., M. V. Swamy, G. Seenayya, and G. Reddy.** 2001. *Clostridium lentocellum* SG6--a potential organism for fermentation of cellulose to acetic acid. Bioresour. Technol. **80:**171-7.

18. **Tatusov, R. L., N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin, E. V. Koonin, D. M. Krylov, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, B. S. Rao, S. Smirnov, A. V. Sverdlov, S. Vasudevan, Y. I. Wolf, J. J. Yin, and D. A. Natale.** 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4:**41.

19. **Zerbino, D. R., and E. Birney.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18:**821-9.

APPENDIX 2

SUPPLEMENTARY TABLES

Table A.2.1 Genomes examined in Chapter 1

ORGANISM NAME	ACCESSION NO.
Acetohalobium arabaticum DSM 5501	NC_014378
Acidaminococcus fermentans DSM 20731	NC_013740
Alicyclobacillus acidocaldarius DSM 446	NC_013205
Alkaliphilus metalliredigens QYMF	NC_009633
Alkaliphilus oremlandii OhILAs	NC_009922
Ammonifex degensii KC4	NC_013385
Anabaena variabilis ATCC 29413	NC_007413
Anaerococcus prevotii DSM 20548	NC_013171
Anoxybacillus flavithermus WK1	NC_011567
Bacillus amyloliquefaciens FZB42	NC_009725
Bacillus anthracis str. Sterne	NC_005945
Bacillus cellulosilyticus DSM 2522	NC_014829
Bacillus cereus ATCC 14579	NC_004722
Bacillus clausii KSM-K16	NC_006582
Bacillus halodurans C-125	NC_002570
Bacillus licheniformis ATCC 14580	NC_006270
Bacillus megaterium QM B1551	NC_014019
Bacillus pseudofirmus OF4	NC_013791
Bacillus pumilus SAFR-032	NC_009848
Bacillus selenitireducens MLS10	NC_014219
Bacillus subtilis subsp. *subtilis* str. 168	NC_000964
Bacillus thuringiensis str. Al Hakam	NC_008600
Bacillus tusciae DSM 2912	NC_014098
Bacillus weihenstephanensis KBAB4	NC_010184
Brevibacillus brevis NBRC 100599	NC_012491
Butyrivibrio proteoclasticus B316	NC_014387, NC_014388
Caldicellulosiruptor bescii DSM 6725	NC_012034
Caldicellulosiruptor hydrothermalis 108	NC_014652
Caldicellulosiruptor kristjanssonii 177R1B	NC_014721
Caldicellulosiruptor kronotskyensis 2002	NC_014720
Caldicellulosiruptor obsidiansis OB47	NC_014392
Caldicellulosiruptor owensensis OL	NC_014657
Caldicellulosiruptor saccharolyticus DSM 8903	NC_009437
Candidatus Desulforudis audaxviator MP104C	NC_010424
Carboxydothermus hydrogenoformans Z-2901	NC_007503
Clostridiales genomosp. BVAB3 str. UPII9-5	NC_013895
Clostridium acetobutylicum ATCC 824	NC_003030
Clostridium beijerinckii NCIMB 8052	NC_009617
Clostridium botulinum A str. ATCC 3502	NC_009495
Clostridium cellulolyticum H10	NC_011898
Clostridium cellulovorans 743B	NC_014393
Clostridium difficile 630	NC_009089
Clostridium kluyveri DSM 555	NC_009706
Clostridium ljungdahlii DSM 13528	NC_014328
Clostridium novyi NT	NC_008593
Clostridium perfringens str. 13	NC_003366
Clostridium phytofermentans ISDg	NC_010001
Clostridium saccharolyticum WM1	NC_014376
Clostridium sticklandii DSM 519	NC_014614

Table A.2.1 (Continued)

ORGANISM NAME	ACCESSION NO.
Clostridium tetani E88	NC_004557
Clostridium thermocellum ATCC 27405	NC_009012
Coprothermobacter proteolyticus DSM 5265	NC_011295
Desulfitobacterium hafniense Y51	NC_007907
Desulfotomaculum acetoxidans DSM 771	NC_013216
Desulfotomaculum reducens MI-1	NC_009253
Enterococcus faecalis V583	NC_004668
Ethanoligenens harbinense YUAN-3	NC_014828
Eubacterium eligens ATCC 27750	NC_012778
Eubacterium limosum KIST612	NC_014624
Eubacterium rectale ATCC 33656	NC_012781
Exiguobacterium sp. AT1b	NC_012673
Exiguobacterium sibiricum 255-15	NC_010556
Finegoldia magna ATCC 29328	NC_010376
Geobacillus sp. C56-T3	NC_014206
Geobacillus kaustophilus HTA426	NC_006510
Geobacillus thermodenitrificans NG80-2	NC_009328
Geobacillus sp. WCH70	NC_012793
Geobacillus sp. Y4.1MC1	NC_014650
Geobacillus sp. Y412MC52	NC_014915
Geobacillus sp. Y412MC61	NC_013411
Halothermothrix orenii H 168	NC_011899
Heliobacterium modesticaldum Ice1	NC_010337
Lactobacillus acidophilus NCFM	NC_006814
Lactobacillus amylovorus GRL 1112	NC_014724
Lactobacillus brevis ATCC 367	NC_008497
Lactobacillus casei ATCC 334	NC_008526
Lactobacillus crispatus ST1	NC_014106
Lactobacillus delbrueckii subsp. *bulgaricus* ATCC 11842	NC_008054
Lactobacillus fermentum IFO 3956	NC_010610
Lactobacillus gasseri ATCC 33323	NC_008530
Lactobacillus helveticus DPC 4571	NC_010080
Lactobacillus johnsonii NCC 533	NC_005362
Lactobacillus plantarum WCFS1	NC_004567
Lactobacillus reuteri DSM 20016	NC_009513
Lactobacillus rhamnosus GG	NC_013198
Lactobacillus sakei subsp. *sakei* 23K	NC_007576
Lactobacillus salivarius UCC118	NC_007929
Lactococcus lactis subsp. *cremoris* MG1363	NC_009004
Leuconostoc citreum KM20	NC_010471
Leuconostoc gasicomitatum LMG 18811	NC_014319
Leuconostoc kimchii IMSNU 11154	NC_014136
Leuconostoc mesenteroides subsp. *mesenteroides* ATCC 8293	NC_008531
Listeria innocua Clip11262	NC_003212
Listeria monocytogenes EGD-e	NC_003210
Listeria seeligeri serovar 1/2b str. SLCC3954	NC_013891
Listeria welshimeri serovar 6b str. SLCC5334	NC_008555
Lysinibacillus sphaericus C3-41	NC_010382

Table A.2.1 (Continued)

ORGANISM NAME	ACCESSION NO.
Macrococcus caseolyticus JCSC5402	NC_011999
Moorella thermoacetica ATCC 39073	NC_007644
Myxococcus xanthus DK 1622	NC_008095
Natranaerobius thermophilus JW/NM-WN-LF	NC_010718
Nostoc punctiforme PCC 73102	NC_010628
Oceanobacillus iheyensis HTE831	NC_004193
Oenococcus oeni PSU-1	NC_008528
Paenibacillus sp. JDR-2	NC_012914
Paenibacillus polymyxa E681	NC_014483
Paenibacillus sp. Y412MC10	NC_013406
Pediococcus pentosaceus ATCC 25745	NC_008525
Pelotomaculum thermopropionicum SI	NC_009454
Ruminococcus albus 7	NC_014833
Staphylococcus aureus subsp. *aureus* NCTC 8325	NC_007795
Staphylococcus carnosus subsp. *carnosus* TM300	NC_012121
Staphylococcus epidermidis RP62A	NC_002976
Staphylococcus haemolyticus JCSC1435	NC_007168
Staphylococcus lugdunensis HKU09-01	NC_013893
Staphylococcus pseudintermedius HKU10-03	NC_014925
Staphylococcus saprophyticus subsp. *saprophyticus* ATCC 15305	NC_007350
Streptococcus agalactiae A909	NC_007432
Streptococcus dysgalactiae subsp. *equisimilis* GGS 124	NC_012891
Streptococcus equi subsp. *equi* 4047	NC_012471
Streptococcus gallolyticus UCN34	NC_013798
Streptococcus gordonii str. Challis substr. CH1	NC_009785
Streptococcus mitis B6	NC_013853
Streptococcus mutans UA159	NC_004350
Streptococcus pneumoniae R6	NC_003098
Streptococcus pyogenes M1 GAS	NC_002737
Streptococcus sanguinis SK36	NC_009009
Streptococcus suis 05ZYH33	NC_009442
Streptococcus thermophilus LMG 18311	NC_006448
Streptococcus uberis 0140J	NC_012004
Streptomyces coelicolor A3(2)	NC_003888
Symbiobacterium thermophilum IAM 14863	NC_006177
Syntrophomonas wolfei subsp. *wolfei* str. Goettingen	NC_008346
Syntrophothermus lipocalidus DSM 12680	NC_014220
Thermincola potens JR	NC_014152
Thermoanaerobacter brockii subsp. *finnii* Ako-1	NC_014964
Thermoanaerobacter italicus Ab9	NC_013921
Thermoanaerobacter mathranii subsp. *mathranii* str. A3	NC_014209
Thermoanaerobacter pseudethanolicus ATCC 33223	NC_010321
Thermoanaerobacter tengcongensis MB4	NC_003869
Thermoanaerobacter sp. X513	NC_014538
Thermoanaerobacter sp. X514	NC_010320
Thermoanaerobacterium thermosaccharolyticum DSM 571	NC_014410
Thermosediminibacter oceani DSM 16646	NC_014377
Veillonella parvula DSM 2008	NC_013520

Table A.2.2. BLAST search results from Chapter 1. Core sporulation-specific genes were searched for in all genomes of sporulating Firmicutes plus those of the potential spore-formers listed in Chapter 1.

ORGANISM	AprX	BofA	CotI	CotJB	CotJC	CotS	CsfB	CwlD	CwlJ	DacB	DacF	GerAA	GerAB
Clostridium sticklandii	0	0	0	0	0	0	0	0	0	0	36	0	0
Alkaliphilus metalliredigens	25	0	0	0	0	0	0	0	0	30	31	39	25
Clostridium difficile	36	0	0	32	51	0	0	37	29	37	44	0	0
Clostridium phytofermentans	42	0	0	0	0	20	0	37	0	32	44	35	0
Clostridium saccharolyticum	41	0	0	32	52	20	0	38	0	33	42	26	0
Acetohalobium arabaticum	33	0	0	39	57	0	0	0	0	0	44	37	23
Clostridium botulinum	29	30	27	43	53	22	0	38	33	0	45	39	29
Thermosediminibacter oceani	0	33	0	34	62	0	0	34	32	33	44	40	27
Alkaliphilus oremlandii	0	0	0	45	61	0	0	39	0	43	47	43	29
Natranaerobius thermophilus	40	32	0	0	35	0	0	42	0	32	35	37	28
Syntrophomonas wolfei	29	30	0	0	0	0	0	43	0	34	42	34	0
Syntrophothermus lipocalidus	0	33	0	0	0	0	0	43	0	0	42	38	23
Eubacterium eligens	36	0	0	29	52	0	0	40	0	0	0	26	0
Eubacterium rectale	35	0	0	0	0	21	0	34	0	0	0	37	0
Caldicellulosiruptor bescii	28	0	26	0	0	26	0	36	0	33	49	41	28
Caldicellulosiruptor kronotskyensis	0	0	27	0	0	26	0	36	0	33	49	41	27
Caldicellulosiruptor hydrothermalis	0	0	26	0	0	26	0	32	0	39	49	41	26
Caldicellulosiruptor kristjanssonii	28	26	28	0	0	26	0	33	0	33	46	41	26
Caldicellulosiruptor saccharolyticus	0	0	26	0	0	25	0	37	0	38	44	38	27
Caldicellulosiruptor obsidiansis	29	0	27	0	0	26	0	33	0	40	50	41	27
Caldicellulosiruptor owensensis	0	27	26	0	0	27	0	33	0	40	50	40	27
Ruminococcus albus	0	0	0	31	50	0	0	37	0	44	29	30	0
Thermoanaerobacter tengcongensis	45	32	0	40	58	0	40	35	0	37	47	41	26
Thermoanaerobacterium thermosaccharolyticum	47	38	0	42	62	0	34	41	0	38	45	40	26
Thermoanaerobacter sp. X513	48	31	0	40	60	0	36	38	0	39	46	40	26
Thermoanaerobacter sp. X514	48	31	0	40	60	0	36	40	0	39	46	40	26
Thermoanaerobacter mathranii	45	0	0	40	60	0	36	38	0	40	45	40	26
Thermoanaerobacter italicus	45	0	0	40	60	0	36	38	0	40	46	40	26
Thermoanaerobacter brockii	48	31	0	40	60	0	36	38	0	39	46	40	26
Thermoanaerobacter pseudethanolicus	48	31	0	40	60	0	36	40	0	39	46	40	26
Lysinibacillus sphaericus	0	0	0	52	87	0	0	41	58	40	31	0	38
Clostridium novyi	30	34	26	40	54	24	0	40	0	0	42	37	28
Halothermothrix orenii	0	0	0	40	54	0	0	0	0	0	46	41	29
Clostridium thermocellum	28	23	34	40	60	28	0	37	30	42	43	43	25
Clostridium kluyveri	33	0	26	35	60	25	0	39	0	38	41	41	0
Clostridium ljungdahlii	30	0	26	36	60	25	0	36	0	37	43	41	0
Bacillus subtilis	100	100	100	100	100	100	100	100	100	100	100	100	100
Clostridium cellulolyticum	0	0	28	36	0	28	0	35	33	41	43	37	22
Ethanoligenens harbinense	0	30	0	37	50	0	0	33	0	0	42	32	27
Clostridium beijerinckii	33	0	26	33	57	25	36	34	0	38	45	23	0
Clostridium perfringens	27	38	28	0	0	25	42	38	0	41	45	32	0
Clostridium cellulovorans	0	0	27	44	59	23	0	37	0	37	43	44	20
Clostridium acetobutylicum	27	0	28	45	53	24	0	36	0	36	45	38	22
Clostridium tetani	30	0	28	0	59	23	0	37	0	37	44	39	27
Paenibacillus polymyxa	34	0	25	0	79	0	40	54	59	41	50	0	35
Paenibacillus sp. JDR-2	32	26	0	52	80	0	34	55	66	44	56	41	33
Paenibacillus sp. Y412MC10	33	24	23	51	82	20	40	55	62	42	53	50	34
Symbiobacterium thermophilum	0	0	32	0	0	26	0	41	49	0	41	40	0
Heliobacterium modesticaldum	32	31	23	36	61	26	0	32	57	38	40	39	22
Alicyclobacillus acidocaldarius	0	0	0	54	84	20	0	40	0	0	50	0	20
Bacillus tusciae	50	0	0	45	86	23	0	31	0	43	51	40	24
Candidatus Desulforudis audaxviator	0	0	0	0	0	0	0	40	0	32	42	0	0
Ammonifex degensii	0	0	0	40	51	0	0	43	0	37	38	35	20
Moorella thermoacetica	35	0	0	37	60	0	0	40	0	36	38	39	24
Thermincola potens	48	37	0	39	59	0	26	42	0	34	41	43	26
Desulfitobacterium hafniense	0	0	0	0	0	0	0	39	0	37	40	39	23
Desulfotomaculum reducens	34	0	0	44	57	0	0	38	0	0	46	43	24
Pelotomaculum thermopropionicum	0	0	0	33	56	0	35	44	0	37	43	39	24
Carboxydothermus hydrogenoformans	0	42	0	0	35	0	0	42	32	0	38	38	20
Desulfotomaculum acetoxidans	0	34	35	33	36	30	38	39	0	35	43	0	23
Oceanobacillus iheyensis	53	42	0	0	40	0	48	51	60	44	53	41	23
Geobacillus sp. WCH70	0	45	20	56	88	0	46	65	58	53	67	44	0
Anoxybacillus flavithermus	0	40	35	0	0	29	41	64	65	55	64	37	26
Bacillus cellulosilyticus	0	48	26	0	0	23	0	55	51	52	61	38	22
Bacillus clausii	35	36	22	0	0	25	0	55	60	56	57	55	33
Geobacillus thermodenitrificans	56	43	54	56	87	43	0	62	59	53	68	39	21
Geobacillus kaustophilus	56	44	35	56	86	27	46	60	54	56	68	39	0
Geobacillus sp. Y4.1MC1	59	42	36	61	89	28	44	63	60	55	67	43	21
Geobacillus sp. C56-T3	57	44	35	56	86	27	46	59	54	56	68	39	0
Geobacillus sp. Y412MC52	57	44	35	56	86	27	46	60	57	56	68	38	0
Geobacillus sp. Y412MC61	57	44	35	56	86	27	46	60	57	56	68	38	0
Bacillus amyloliquefaciens	81	75	0	73	96	0	64	86	87	77	85	78	68
Bacillus pumilus	69	64	0	0	32	0	66	75	88	68	74	65	51
Brevibacillus brevis	0	40	0	54	83	0	34	44	72	48	54	44	34
Bacillus weihenstephanensis	35	41	69	54	88	59	38	59	83	51	61	40	25
Bacillus cereus	34	43	0	56	89	0	34	59	61	48	61	40	37
Bacillus anthracis	34	43	0	60	89	0	38	59	62	47	62	48	29
Bacillus thuringiensis	31	42	0	60	89	0	0	59	62	48	62	40	25
Bacillus halodurans	69	45	38	0	33	28	40	56	58	56	59	45	25
Bacillus pseudofirmus	34	52	23	0	31	23	28	57	53	52	59	39	26
Bacillus licheniformis	71	62	0	55	89	22	68	76	83	73	78	67	59
Bacillus megaterium	33	49	0	56	91	0	33	58	56	57	64	52	44

Table A.2.2 (Continued)

ORGANISM	GerAC	GerBA	GerBB	GerBC	GerE	GerKA	GerKB	GerKC	Gpr	Jag	KamA	KapD	KipR	LonB
C. sticklandii	0	0	0	0	0	0	0	0	0	34	60	0	25	0
A. metalliredigens	24	38	25	21	0	41	32	23	0	0	0	0	29	60
C. difficile	0	0	0	0	0	0	0	0	41	35	49	0	26	0
C. phytofermentans	0	35	0	0	0	37	0	0	40	39	54	0	0	0
C. saccharolyticum	0	25	0	0	0	27	0	0	38	39	46	0	29	0
A. arabaticum	25	37	20	25	0	45	28	27	38	38	51	0	29	63
C. botulinum	24	41	22	0	0	36	26	23	41	37	0	21	0	59
T. oceani	25	39	25	24	0	43	27	26	46	42	0	33	29	62
A. oremlandii	23	43	25	21	0	45	23	24	42	0	59	0	0	57
N. thermophilus	24	35	26	23	0	40	28	26	42	42	58	0	28	60
S. wolfei	22	37	20	22	0	44	30	25	33	39	45	0	0	57
S. lipocalidus	21	40	21	23	0	42	26	24	38	41	47	0	0	59
E. eligens	0	24	0	0	0	25	0	0	43	0	0	0	0	0
E. rectale	0	39	0	0	0	31	0	0	44	34	0	0	0	0
C. bescii	24	38	22	23	0	47	25	24	41	40	47	0	30	0
C. kronotskyensis	24	37	22	24	0	47	25	24	41	40	49	0	30	0
C. hydrothermalis	24	38	23	22	0	48	24	22	42	39	48	0	36	0
C. kristjanssonii	23	38	23	22	0	47	24	23	42	39	47	0	29	0
C. saccharolyticus	22	38	23	22	0	47	22	24	41	39	46	0	31	0
C. obsidiansis	23	39	22	22	0	48	24	22	41	40	49	0	30	0
C. owensensis	23	38	22	22	0	47	25	24	42	38	47	0	30	0
R. albus	0	30	0	0	0	34	0	21	38	30	0	0	0	0
T. tengcongensis	23	40	26	26	0	50	25	25	45	40	60	0	0	0
T. thermosaccharolyticum	23	39	23	26	0	48	32	26	41	40	0	0	0	0
Thermoanaerobacter sp. X513	23	39	25	26	0	46	29	25	43	40	0	0	0	0
Thermoanaerobacter sp. X514	23	39	25	26	0	46	29	25	43	40	0	0	0	0
T. mathranii	23	39	25	25	0	46	25	26	43	40	0	0	0	0
T. italicus	23	39	25	25	0	46	29	26	43	40	0	0	0	0
T. brockii	23	39	25	27	0	46	29	25	43	40	0	0	33	0
T. pseudethanolicus	23	39	25	27	0	46	29	25	43	40	0	0	33	0
L. sphaericus	29	42	35	29	65	51	0	31	36	35	73	0	37	0
C. novyi	25	42	0	25	0	42	24	21	42	40	0	0	29	59
H. orenii	23	41	26	24	0	40	24	22	37	40	0	0	29	65
C. thermocellum	24	39	24	23	0	38	0	26	46	42	0	29	0	59
C. kluyveri	0	42	0	0	0	42	24	24	44	37	0	0	21	41
C. ljungdahlii	0	40	0	28	0	46	28	28	41	39	0	0	27	43
B. subtilis	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C. cellulolyticum	22	36	22	24	0	40	28	23	43	41	0	27	0	60
E. harbinense	22	32	0	26	0	37	28	27	43	35	0	0	0	0
C. beijerinckii	0	26	0	0	0	28	0	0	41	36	0	0	30	59
C. perfringens	0	32	0	21	0	33	0	20	41	42	0	0	0	59
C. cellulovorans	22	40	20	0	0	39	0	23	41	43	42	0	26	57
C. acetobutylicum	23	40	21	27	0	46	31	27	43	40	0	0	0	53
C. tetani	21	45	22	22	0	41	22	24	43	39	57	0	0	59
P. polymyxa	28	41	33	27	75	44	32	30	47	31	0	27	30	63
Paenibacillus sp. JDR-2	28	40	35	26	75	46	28	29	49	38	27	0	31	64
Paenibacillus sp. Y412MC10	25	41	33	28	75	47	32	30	48	32	0	0	28	64
S. thermophilum	0	37	21	20	0	44	24	23	40	38	59	0	32	59
H. modesticaldum	21	34	20	24	0	41	27	23	37	0	0	0	0	61
A. acidocaldarius	20	0	22	23	77	45	24	23	41	33	0	0	30	64
B. tusciae	0	37	26	21	77	42	26	24	47	0	0	49	30	55
Ca. D. audaxviator	0	35	0	0	0	35	23	0	40	39	54	0	0	60
A. degensii	25	35	20	24	0	42	27	26	40	40	47	0	28	60
M. thermoacetica	21	41	24	26	0	46	31	27	37	35	46	0	35	60
T. potens	25	40	24	26	0	46	27	26	40	36	47	0	30	65
D. hafniense	22	37	22	27	0	44	27	27	38	36	46	34	27	63
D. reducens	24	37	21	22	0	40	26	23	42	39	60	0	27	66
P. thermopropionicum	24	40	20	21	0	46	26	26	43	39	45	0	29	64
C. hydrogenoformans	21	38	23	24	0	43	25	27	38	37	49	0	35	54
D. acetoxidans	22	43	20	24	0	46	25	25	41	41	53	0	0	63
O. iheyensis	24	38	0	25	87	38	21	23	60	47	0	24	31	0
Geobacillus sp. WCH70	23	39	0	25	91	45	27	30	67	48	30	0	32	77
A. flavithermus	0	37	24	20	94	55	47	40	63	40	0	0	31	74
B. cellulosilyticus	22	38	22	21	86	48	27	22	63	42	31	0	0	69
B. clausii	32	37	27	30	81	44	23	27	60	39	0	0	32	66
G. thermodenitrificans	22	40	20	26	93	57	47	38	65	47	0	56	51	77
G. kaustophilus	24	40	20	25	93	59	45	39	64	44	0	56	51	77
Geobacillus sp. Y4.1MC1	24	40	22	24	93	58	47	43	66	50	30	58	51	76
Geobacillus sp. C56-T3	23	41	20	25	93	58	46	39	64	44	0	55	51	77
Geobacillus sp. Y412MC52	0	40	20	23	93	58	45	40	64	44	0	55	51	77
Geobacillus sp. Y412MC61	0	40	20	23	93	58	45	40	64	44	0	55	51	77
B. amyloliquefaciens	57	70	68	66	98	72	73	68	85	69	89	85	79	92
B. pumilus	44	56	47	47	100	41	26	21	78	49	83	66	67	82
B. brevis	31	45	34	29	88	50	29	30	56	37	71	34	48	62
B. weihenstephanensis	25	37	24	24	95	37	23	22	69	47	80	56	53	74
B. cereus	32	38	33	30	95	38	25	22	68	47	78	57	52	75
B. anthracis	25	45	28	26	95	44	0	23	68	48	79	56	52	75
B. thuringiensis	25	37	24	25	95	37	21	23	68	48	79	56	52	75
B. halodurans	23	41	24	24	85	47	40	34	65	42	76	0	57	70
B. pseudofirmus	25	43	25	29	83	42	22	26	64	43	79	0	52	71
B. licheniformis	47	48	34	32	100	64	58	52	76	60	84	77	71	86
B. megaterium	36	47	34	33	98	60	54	45	67	40	80	58	62	76

Table A.2.2 (Continued)

ORGANISM	MmgC	NtdA	ParB	PbpG	PbpI	PdaA	PrkA	SigE	SigF	SigG	SigK	SleB	SplB	SpmA	SpmB
C. sticklandii	52	36	42	0	23	0	0	0	37	34	0	0	0	0	0
A. metalliredigens	0	0	0	32	26	29	0	49	49	51	58	0	47	41	45
C. difficile	54	0	43	27	25	43	0	67	51	68	55	43	45	32	32
C. phytofermentans	0	0	41	0	0	45	0	65	53	65	53	0	33	43	37
C. saccharolyticum	53	34	40	29	28	43	0	67	50	64	45	0	33	37	41
A. arabaticum	0	37	43	34	25	32	0	71	52	70	54	0	43	44	53
C. botulinum	0	32	42	30	28	43	0	71	50	73	56	0	44	39	47
T. oceani	0	34	47	36	25	32	0	71	50	74	61	53	51	43	50
A. oremlandii	55	31	0	30	25	31	0	72	50	74	62	0	43	48	53
N. thermophilus	55	0	44	35	26	0	0	73	51	74	46	40	49	49	52
S. wolfei	48	32	40	37	23	38	0	66	51	75	58	0	31	48	56
S. lipocalidus	49	32	40	37	27	0	0	72	51	74	54	0	0	46	56
E. eligens	0	24	42	30	24	30	0	66	49	54	46	0	0	0	0
E. rectale	51	33	42	27	24	32	0	65	45	60	52	28	20	0	0
C. bescii	47	34	41	33	24	29	0	60	48	61	51	0	21	32	34
C. kronotskyensis	48	36	41	33	24	29	0	60	48	60	51	0	21	32	33
C. hydrothermalis	47	36	40	33	24	30	0	60	48	60	52	0	21	34	34
C. kristjanssonii	47	36	40	33	24	29	0	60	48	61	52	0	22	34	32
C. saccharolyticus	46	35	43	33	24	27	0	61	50	62	53	49	22	35	31
C. obsidiansis	47	37	39	33	24	28	0	59	49	60	51	0	22	33	30
C. owensensis	47	36	40	33	24	29	0	61	49	60	52	0	22	32	32
R. albus	0	26	38	26	0	0	0	63	46	55	47	0	0	0	0
T. tengcongensis	57	37	45	32	26	46	0	72	50	73	56	0	22	42	46
T. thermosaccharolyticum	58	0	48	34	27	45	0	70	50	75	57	0	22	46	53
Thermoanaerobacter sp. X513	0	0	48	34	27	48	0	70	50	74	57	0	23	43	50
Thermoanaerobacter sp. X514	0	0	48	34	27	48	0	70	50	74	57	0	23	43	50
T. mathranii	0	30	48	36	27	46	0	69	51	74	58	0	22	42	50
T. italicus	0	0	48	36	27	46	0	69	51	74	57	0	22	42	50
T. brockii	26	0	47	34	27	46	0	70	50	74	57	0	23	43	50
T. pseudethanolicus	26	0	47	34	27	46	0	70	50	74	57	0	23	43	50
L. sphaericus	65	26	57	45	23	47	83	61	46	60	50	54	82	0	0
C. novyi	52	32	41	29	25	44	0	69	52	75	57	0	45	36	43
H. orenii	0	27	43	34	24	30	0	65	47	71	55	0	48	47	46
C. thermocellum	0	33	47	31	28	42	40	69	55	74	58	0	0	42	50
C. kluyveri	0	32	43	31	26	45	38	69	50	75	55	0	45	37	53
C. ljungdahlii	42	33	43	31	26	44	38	70	49	74	53	0	43	40	48
B. subtilis	100	100	100	100	100	100	100	100	100	100	47	100	100	100	100
C. cellulolyticum	36	0	45	31	27	42	40	74	54	73	46	0	43	44	46
E. harbinense	38	33	42	29	24	30	38	59	50	66	52	0	30	37	41
C. beijerinckii	0	31	49	30	23	43	40	71	50	74	55	40	43	37	40
C. perfringens	54	0	46	26	24	41	41	71	52	75	54	38	47	40	42
C. cellulovorans	51	29	44	28	27	0	39	70	48	72	55	0	0	35	43
C. acetobutylicum	52	27	46	30	26	41	40	74	49	73	57	0	42	36	45
C. tetani	0	30	44	30	26	43	38	72	48	74	47	0	44	38	51
P. polymyxa	0	0	49	40	30	50	73	77	73	79	71	44	61	50	58
Paenibacillus sp. JDR-2	53	34	49	45	30	50	73	78	74	79	73	48	64	53	55
Paenibacillus sp. Y412MC10	0	35	47	41	31	51	72	80	72	79	71	42	59	56	58
S. thermophilum	54	0	47	31	24	34	69	70	54	73	56	0	0	42	53
H. modesticaldum	55	33	36	34	21	0	73	76	55	76	61	0	49	50	50
A. acidocaldarius	50	0	51	37	25	34	72	77	65	77	57	35	0	47	48
B. tusciae	56	31	51	40	30	0	75	71	68	76	70	0	62	49	54
Ca. D. audaxviator	0	36	38	34	26	0	66	74	51	74	58	0	46	47	51
A. degensii	0	27	38	35	23	34	61	69	49	70	44	0	0	47	56
M. thermoacetica	0	35	44	0	24	33	71	71	54	73	64	0	48	49	55
T. potens	0	0	45	40	28	37	71	78	56	76	64	0	50	48	50
D. hafniense	59	37	46	33	24	36	69	76	55	73	55	0	33	49	58
D. reducens	54	30	43	40	25	41	73	78	54	75	48	0	51	48	50
P. thermopropionicum	48	37	39	39	26	49	75	71	53	74	46	0	49	48	51
C. hydrogenoformans	58	0	45	36	24	30	72	71	51	75	57	53	0	48	52
D. acetoxidans	54	35	39	38	27	0	72	74	51	74	63	57	45	49	50
O. iheyensis	65	28	55	46	26	53	84	84	72	82	73	56	75	56	65
Geobacillus sp. WCH70	68	27	66	57	48	61	88	83	82	89	84	62	0	62	74
A. flavithermus	68	0	67	54	47	58	90	85	80	89	77	62	73	68	73
B. cellulosilyticus	61	30	62	47	25	52	83	84	75	84	75	0	78	70	70
B. clausii	60	0	58	47	26	49	85	85	77	83	74	57	73	61	63
G. thermodenitrificans	69	0	65	56	49	59	88	82	80	87	81	61	71	62	73
G. kaustophilus	69	37	63	56	49	58	87	82	82	89	81	61	0	64	72
Geobacillus sp. Y4.1MC1	68	37	66	58	48	61	90	83	82	87	83	60	0	62	74
Geobacillus sp. C56-T3	69	38	63	56	49	59	87	82	82	89	80	60	72	64	73
Geobacillus sp. Y412MC52	69	37	63	56	49	59	87	82	82	89	81	60	72	64	72
Geobacillus sp. Y412MC61	69	37	63	56	49	59	87	82	82	89	81	60	72	64	72
B. amyloliquefaciens	77	33	91	79	79	88	97	97	96	98	96	76	94	93	94
B. pumilus	56	58	81	68	61	75	93	93	88	97	90	68	90	81	84
B. brevis	70	28	56	43	34	50	76	80	75	79	47	38	65	53	57
B. weihenstephanensis	67	31	61	51	44	52	87	87	79	88	85	54	75	65	72
B. cereus	67	26	62	51	43	53	87	87	80	89	85	53	74	67	73
B. anthracis	66	59	61	51	45	53	88	87	80	88	85	47	74	66	72
B. thuringiensis	66	58	61	51	45	53	88	87	80	88	85	47	74	66	72
B. halodurans	67	0	61	50	26	50	85	85	76	84	76	56	75	66	70
B. pseudofirmus	66	28	59	50	27	48	86	86	80	85	74	52	73	64	72
B. licheniformis	57	53	82	65	71	75	94	96	91	95	96	69	89	87	91
B. megaterium	63	0	63	53	49	62	88	87	83	91	87	60	85	69	74

Table A.2.2 (Continued)

	Spo0A	Spo0F	SpoIIAA	SpoIIAB	SpoIID	SpoIIE	SpoIIGA	SpoIIIAA	SpoIIIAB	SpoIIIAC	SpoIIIAD	SpoIIIAE
C. sticklandii	30	38	28	28	28	0	0	0	0	0	0	0
A. metalliredigens	0	37	37	56	34	30	0	41	33	40	48	41
C. difficile	56	39	43	55	35	26	0	38	26	37	39	25
C. phytofermentans	55	40	35	51	0	28	21	43	25	0	37	22
C. saccharolyticum	49	37	36	47	29	23	20	40	27	40	36	22
A. arabaticum	60	0	40	43	37	27	22	37	32	37	46	40
C. botulinum	57	35	35	54	35	25	23	43	28	41	40	33
T. oceani	56	37	41	56	40	27	28	43	35	43	42	38
A. oremlandii	55	0	35	59	0	0	27	40	26	40	46	37
N. thermophilus	53	0	34	53	33	26	23	43	30	37	49	41
S. wolfei	52	39	38	43	35	27	22	45	32	37	37	33
S. lipocalidus	52	47	33	43	32	29	24	45	27	38	34	29
E. eligens	46	0	0	51	0	0	0	39	0	33	0	24
E. rectale	0	34	37	45	0	0	0	40	0	30	33	23
C. bescii	47	37	36	53	36	23	23	34	0	33	30	22
C. kronotskyensis	48	37	36	53	36	23	23	33	0	33	30	22
C. hydrothermalis	47	37	36	53	38	23	22	34	0	33	31	22
C. kristjanssonii	48	37	36	54	37	23	22	33	0	33	30	21
C. saccharolyticus	45	38	34	56	35	24	24	38	0	0	33	25
C. obsidiansis	47	37	33	55	37	23	23	34	0	0	30	22
C. owensensis	48	37	37	54	38	23	24	34	0	0	31	22
R. albus	0	0	34	47	30	24	21	32	0	40	27	0
T. tengcongensis	51	35	37	52	38	28	26	45	35	38	35	34
T. thermosaccharolyticum	51	0	34	48	34	28	26	41	31	41	44	32
Thermoanaerobacter sp. X513	53	0	33	53	37	28	25	44	33	37	39	34
Thermoanaerobacter sp. X514	53	0	33	53	37	28	25	44	33	37	39	34
T. mathranii	53	0	32	53	36	28	25	45	34	37	39	34
T. italicus	53	0	32	53	36	28	25	45	34	37	39	34
T. brockii	53	35	33	53	37	28	26	44	33	37	39	32
T. pseudethanolicus	53	35	33	53	37	28	26	44	33	37	39	32
L. sphaericus	59	61	37	58	36	23	0	0	0	33	0	20
C. novyi	58	36	37	52	35	26	20	45	25	42	43	33
H. orenii	52	41	37	41	34	25	23	40	23	40	46	41
C. thermocellum	58	36	40	51	36	28	26	43	28	34	47	35
C. kluyveri	55	35	37	55	37	25	25	45	31	38	39	35
C. ljungdahlii	57	36	37	54	38	25	24	45	31	37	40	36
B. subtilis	100	100	100	100	100	100	100	100	100	100	100	100
C. cellulolyticum	53	35	32	60	35	26	26	40	34	37	44	36
E. harbinense	41	34	31	52	35	21	25	39	28	38	25	20
C. beijerinckii	57	35	35	51	37	26	0	41	26	35	40	30
C. perfringens	75	0	44	57	40	24	23	40	26	49	35	25
C. cellulovorans	56	36	35	61	33	25	21	39	24	40	36	28
C. acetobutylicum	53	39	35	53	36	24	21	41	28	35	39	35
C. tetani	55	36	37	53	36	24	25	42	28	0	0	34
P. polymyxa	68	72	52	59	41	44	34	49	37	55	56	51
Paenibacillus sp. JDR-2	68	68	52	58	45	44	31	47	43	58	58	54
Paenibacillus sp. Y412MC10	69	68	53	59	0	45	31	50	44	59	61	52
S. thermophilum	43	33	37	41	35	0	0	48	33	50	40	38
H. modesticaldum	56	34	35	41	33	0	25	53	28	42	42	41
A. acidocaldarius	62	59	38	55	28	32	24	44	26	38	40	35
B. tusciae	61	58	42	65	41	38	23	47	36	54	50	44
Ca. D. audaxviator	54	42	41	41	36	25	24	42	33	40	46	40
A. degensii	54	34	34	40	38	26	27	39	32	41	44	37
M. thermoacetica	54	38	41	46	43	0	29	42	29	34	47	37
T. potens	54	47	43	47	41	30	25	47	38	41	44	39
D. hafniense	55	42	44	43	42	22	25	43	30	35	39	34
D. reducens	55	0	38	41	41	28	23	41	38	40	50	44
P. thermopropionicum	56	46	39	43	38	28	23	42	39	46	50	40
C. hydrogenoformans	55	0	37	38	35	25	0	44	30	48	50	31
D. acetoxidans	55	45	44	43	35	26	27	45	35	42	50	42
O. iheyensis	70	0	55	67	45	55	36	46	35	57	46	51
Geobacillus sp. WCH70	81	80	73	76	58	59	48	55	56	86	66	65
A. flavithermus	77	76	68	78	53	56	44	55	52	79	65	62
B. cellulosilyticus	74	67	63	71	44	53	39	54	41	68	68	57
B. clausii	74	57	67	66	44	52	38	52	48	66	61	57
G. thermodenitrificans	78	79	73	73	57	58	45	56	53	80	67	60
G. kaustophilus	78	79	74	76	58	57	46	56	53	80	68	59
Geobacillus sp. Y4.1MC1	79	81	75	75	54	58	46	55	57	86	70	63
Geobacillus sp. C56-T3	77	80	74	76	57	57	45	56	53	80	68	59
Geobacillus sp. Y412MC52	78	80	74	76	56	57	45	56	53	80	68	57
Geobacillus sp. Y412MC61	78	80	74	76	56	57	45	56	53	80	68	57
B. amyloliquefaciens	96	95	91	97	77	91	76	77	80	100	95	85
B. pumilus	89	91	84	88	65	79	59	68	64	100	86	84
B. brevis	70	72	60	64	47	47	33	51	41	50	65	54
B. weihenstephanensis	80	77	68	79	52	56	48	61	56	56	76	64
B. cereus	81	77	70	78	52	56	49	61	54	56	76	59
B. anthracis	80	77	70	79	53	56	48	62	54	56	74	64
B. thuringiensis	80	77	70	79	52	56	48	62	54	56	74	64
B. halodurans	76	69	68	73	48	54	44	56	52	70	69	63
B. pseudofirmus	76	65	72	72	49	57	41	57	51	72	73	62
B. licheniformis	93	91	83	88	70	81	69	65	73	95	92	80
B. megaterium	82	79	68	76	53	60	49	56	54	79	74	73

Table A.2.2 (Continued)

ORGANISM	SpoIIIAF	SpoIIIAG	SpoIIIAH	SpoIIID	SpoIIP	SpoIIR	SpoIVA	SpoIVB	SpoIVFB	SpoVAA	SpoVAB	SpoVAC
C. sticklandii	0	0	0	0	0	0	0	0	0	0	0	0
A. metalliredigens	0	0	0	59	0	0	0	0	26	44	43	48
C. difficile	0	32	28	64	0	39	48	37	0	0	0	48
C. phytofermentans	22	30	0	56	0	33	41	33	0	0	0	44
C. saccharolyticum	0	25	0	57	0	36	44	38	0	30	35	38
A. arabaticum	22	0	0	58	27	28	53	47	28	0	0	39
C. botulinum	28	35	0	62	26	39	53	44	0	0	0	44
T. oceani	23	0	30	64	29	0	58	42	0	0	0	52
A. oremlandii	0	0	0	0	0	0	50	42	25	37	40	44
N. thermophilus	24	0	25	55	26	29	56	33	22	0	0	54
S. wolfei	0	0	0	58	28	40	56	39	24	0	0	43
S. lipocalidus	21	0	0	60	30	0	54	38	0	0	0	47
E. eligens	0	0	0	59	0	32	39	32	0	26	28	42
E. rectale	0	31	0	54	24	0	41	35	0	0	30	36
C. bescii	0	23	0	44	0	29	55	40	0	0	0	49
C. kronotskyensis	0	23	0	45	0	28	55	39	0	0	0	50
C. hydrothermalis	0	23	0	44	0	24	54	39	0	0	0	50
C. kristjanssonii	0	23	0	45	0	25	54	40	0	0	0	48
C. saccharolyticus	0	22	0	50	0	0	55	38	0	0	0	46
C. obsidiansis	0	23	0	46	0	25	54	40	0	0	0	50
C. owensensis	0	23	0	43	0	25	54	38	0	0	0	49
R. albus	0	26	0	57	0	0	48	33	27	0	0	39
T. tengcongensis	0	0	26	67	0	35	49	41	27	0	0	49
T. thermosaccharolyticum	0	0	0	67	0	39	50	46	29	0	0	48
Thermoanaerobacter sp. X513	0	31	26	66	0	36	50	41	26	0	0	48
Thermoanaerobacter sp. X514	0	31	26	66	0	36	50	41	26	0	0	48
T. mathranii	0	27	28	66	0	34	50	41	0	0	0	49
T. italicus	0	27	27	66	0	34	50	41	25	0	0	49
T. brockii	0	30	28	67	0	37	50	41	26	0	0	48
T. pseudethanolicus	0	30	28	67	0	37	50	41	26	0	0	48
L. sphaericus	0	0	31	72	28	28	0	27	0	0	0	61
C. novyi	28	32	0	61	0	39	51	43	0	0	0	49
H. orenii	24	0	0	60	0	0	54	43	23	0	0	47
C. thermocellum	24	0	23	60	0	37	56	39	0	0	0	44
C. kluyveri	30	0	0	62	27	38	53	44	0	0	0	58
C. ljungdahlii	25	0	0	62	27	33	52	44	0	0	0	55
B. subtilis	100	100	100	100	100	100	100	100	100	100	100	100
C. cellulolyticum	0	0	0	0	0	34	0	43	0	0	0	46
E. harbinense	0	0	0	0	0	0	52	39	27	0	0	44
C. beijerinckii	0	32	0	56	0	32	49	45	0	0	0	42
C. perfringens	0	33	0	62	0	37	47	46	0	0	0	45
C. cellulovorans	25	0	0	62	0	34	48	40	0	0	0	46
C. acetobutylicum	23	32	0	59	0	34	49	40	0	0	0	45
C. tetani	24	32	21	0	29	0	54	47	0	0	0	46
P. polymyxa	0	32	24	87	35	40	78	57	28	44	45	55
Paenibacillus sp. JDR-2	0	31	28	88	33	37	78	49	29	42	48	55
Paenibacillus sp. Y412MC10	22	31	0	87	36	43	78	48	31	45	50	60
S. thermophilum	0	0	0	60	0	0	53	41	0	0	0	52
H. modesticaldum	24	0	0	60	0	0	57	41	0	0	0	60
A. acidocaldarius	0	0	20	81	0	0	71	34	29	0	0	52
B. tusciae	0	0	20	81	31	0	74	46	26	41	47	54
Ca. D. audaxviator	0	0	0	58	30	40	55	40	25	0	0	45
A. degensii	0	0	0	60	36	0	55	41	0	0	0	44
M. thermoacetica	0	0	0	58	28	39	55	40	0	0	0	40
T. potens	22	31	0	60	28	37	59	46	0	0	0	46
D. hafniense	0	0	0	59	0	0	52	44	28	0	0	50
D. reducens	22	32	0	62	28	42	60	48	25	0	0	58
P. thermopropionicum	0	27	26	60	0	45	58	44	0	0	0	47
C. hydrogenoformans	0	35	0	58	0	37	54	42	0	0	0	47
D. acetoxidans	0	29	0	61	0	36	55	44	0	0	0	58
O. iheyensis	34	0	34	82	44	57	81	63	36	44	53	58
Geobacillus sp. WCH70	42	46	39	94	50	53	89	64	47	57	62	57
A. flavithermus	40	49	43	97	49	55	88	61	45	52	58	57
B. cellulosilyticus	31	45	32	84	38	0	84	58	40	46	56	63
B. clausii	31	45	33	83	42	45	87	51	34	0	0	56
G. thermodenitrificans	42	50	41	94	52	0	89	61	49	59	60	57
G. kaustophilus	41	51	42	94	51	43	89	62	49	56	60	57
Geobacillus sp. Y4.1MC1	42	47	41	94	50	48	89	65	48	55	61	57
Geobacillus sp. C56-T3	41	51	41	94	51	43	89	62	50	56	60	57
Geobacillus sp. Y412MC52	41	51	41	94	51	43	89	62	50	56	62	57
Geobacillus sp. Y412MC61	41	51	41	94	51	43	89	62	50	56	62	57
B. amyloliquefaciens	79	80	82	100	83	77	96	83	77	84	87	90
B. pumilus	50	73	62	97	67	62	94	75	64	64	75	86
B. brevis	0	36	30	87	37	44	75	50	30	48	48	63
B. weihenstephanensis	34	43	43	0	45	0	88	64	49	47	53	71
B. cereus	37	50	42	86	45	50	88	63	48	48	56	71
B. anthracis	37	50	41	0	45	0	88	63	50	47	55	73
B. thuringiensis	37	50	41	0	45	0	88	63	49	48	55	72
B. halodurans	34	44	34	85	41	0	87	55	40	0	0	55
B. pseudofirmus	34	43	33	88	46	50	86	57	36	53	59	61
B. licheniformis	63	69	59	94	71	65	95	80	70	68	72	83
B. megaterium	37	43	47	97	47	54	90	65	48	53	63	78

Table A.2.2 (Continued)

ORGANISM	SpoVAD	spoVAEB	SpoVAF	SpoVB	SpoVFB	SpoVR	SpoVS	SpoVT	SpsF	SpsG	SpsJ	SspA	SspB
C. sticklandii	0	0	0	27	0	0	88	0	0	0	51	0	0
A. metalliredigens	47	48	28	27	0	0	0	61	0	0	28	45	42
C. difficile	46	42	0	27	50	0	82	55	0	0	26	46	48
C. phytofermentans	43	42	34	27	46	0	0	59	0	0	51	39	40
C. saccharolyticum	43	46	33	27	49	0	0	55	0	0	51	41	41
A. arabaticum	51	46	40	30	52	0	89	54	21	0	25	0	31
C. botulinum	47	44	31	30	0	0	93	55	32	0	47	41	40
T. oceani	48	46	30	31	0	0	84	61	0	0	27	34	32
A. oremlandii	46	44	32	27	0	0	93	0	33	23	26	48	49
N. thermophilus	47	57	45	29	53	0	88	59	0	0	29	36	34
S. wolfei	48	39	47	31	47	0	91	62	33	23	32	40	40
S. lipocalidus	46	42	33	30	52	0	93	60	0	0	26	0	36
E. eligens	42	44	35	23	0	0	0	46	0	0	52	40	42
E. rectale	41	44	31	22	46	0	0	36	0	0	25	38	41
C. bescii	45	51	30	27	50	0	68	54	0	0	52	39	37
C. kronotskyensis	45	50	30	22	51	0	68	55	0	0	29	39	37
C. hydrothermalis	45	49	30	27	49	0	68	55	0	0	28	39	37
C. kristjanssonii	44	49	30	27	50	0	68	54	0	0	27	39	37
C. saccharolyticus	44	50	30	23	50	0	68	56	0	0	27	39	37
C. obsidiansis	45	50	31	22	50	0	68	55	0	0	52	39	37
C. owensensis	45	50	29	22	49	0	68	55	0	0	27	39	37
R. albus	42	48	24	21	43	0	66	49	0	0	53	0	0
T. tengcongensis	50	46	29	27	29	0	88	56	0	0	28	0	36
T. thermosaccharolyticum	49	51	31	26	0	0	90	56	0	0	30	38	37
Thermoanaerobacter sp. X513	50	45	30	27	28	0	90	56	0	0	26	0	36
Thermoanaerobacter sp. X514	50	45	30	27	28	0	90	56	0	0	26	0	36
T. mathranii	49	48	29	27	0	0	90	56	33	30	33	38	38
T. italicus	49	48	29	27	0	0	90	56	0	0	52	0	38
T. brockii	50	45	30	27	0	0	90	56	0	0	50	0	36
T. pseudethanolicus	50	45	30	27	0	0	90	56	0	0	50	0	36
L. sphaericus	42	61	46	33	60	74	76	60	0	20	28	65	66
C. novyi	48	40	30	28	0	0	76	59	32	24	28	40	39
H. orenii	49	45	40	32	53	0	90	58	25	0	29	30	30
C. thermocellum	46	41	28	30	53	37	77	59	30	24	38	27	29
C. kluyveri	50	60	30	29	0	36	94	56	0	0	31	42	50
C. ljungdahlii	0	58	31	30	0	37	88	54	0	23	49	41	44
B. subtilis	100	100	100	100	100	100	100	100	100	100	100	100	100
C. cellulolyticum	48	43	30	29	50	37	69	58	0	0	43	35	34
E. harbinense	44	43	26	25	47	38	60	52	0	0	50	40	40
C. beijerinckii	46	45	35	27	0	36	79	62	31	23	48	44	40
C. perfringens	46	45	39	27	0	38	77	63	0	0	50	43	43
C. cellulovorans	47	44	30	29	0	35	89	59	0	0	47	46	48
C. acetobutylicum	47	38	30	29	0	36	89	61	35	22	48	50	50
C. tetani	48	42	31	28	0	30	0	58	31	24	27	0	0
P. polymyxa	51	57	44	39	59	63	90	65	0	0	52	50	46
Paenibacillus sp. JDR-2	55	59	43	40	53	65	91	65	0	0	55	50	50
Paenibacillus sp. Y412MC10	54	56	46	38	58	64	91	66	0	0	53	50	51
S. thermophilum	47	52	31	27	48	50	92	67	0	0	30	41	43
H. modesticaldum	50	64	46	28	52	0	94	60	33	0	28	0	0
A. acidocaldarius	50	64	34	33	56	57	86	65	0	0	24	73	70
B. tusciae	44	62	38	35	62	62	95	65	0	0	53	42	42
Ca. D. audaxviator	50	48	0	31	51	44	90	60	32	25	51	42	41
A. degensii	50	44	46	35	58	43	87	60	0	0	50	41	39
M. thermoacetica	52	42	47	28	56	56	77	61	0	0	55	0	0
T. potens	51	47	48	28	50	61	95	63	0	0	47	34	32
D. hafniense	47	49	48	32	50	47	94	67	0	0	50	55	54
D. reducens	50	61	47	33	55	55	90	63	0	0	50	48	47
P. thermopropionicum	51	44	49	33	53	60	95	62	0	0	52	28	27
C. hydrogenoformans	48	50	31	29	49	54	87	61	0	0	47	38	39
D. acetoxidans	51	60	47	33	53	57	94	65	0	0	51	46	0
O. iheyensis	45	62	58	52	60	74	91	66	0	0	46	74	78
Geobacillus sp. WCH70	65	63	64	64	78	79	95	78	0	0	24	81	81
A. flavithermus	53	61	65	62	76	79	98	77	0	0	27	77	77
B. cellulosilyticus	49	64	61	57	66	76	89	73	36	23	51	77	79
B. clausii	49	61	56	58	64	74	88	70	0	0	51	81	81
G. thermodenitrificans	63	62	65	62	78	80	0	78	0	0	28	77	79
G. kaustophilus	62	63	65	62	79	79	96	78	0	0	27	80	79
Geobacillus sp. Y4.1MC1	65	63	66	63	77	80	96	78	34	26	27	80	80
Geobacillus sp. C56-T3	62	63	65	61	79	80	96	78	0	0	25	80	79
Geobacillus sp. Y412MC52	62	63	65	62	79	80	96	78	38	0	31	80	79
Geobacillus sp. Y412MC61	62	63	65	62	79	80	96	78	38	0	31	80	79
B. amyloliquefaciens	90	91	80	87	83	90	100	89	62	53	84	90	87
B. pumilus	82	84	75	74	84	84	97	85	48	40	72	83	87
B. brevis	58	64	51	41	58	67	96	66	0	0	28	79	78
B. weihenstephanensis	67	72	61	57	79	77	91	73	0	0	51	70	75
B. cereus	67	72	63	57	75	79	91	76	0	0	50	71	78
B. anthracis	66	76	63	56	74	78	91	76	0	0	50	75	78
B. thuringiensis	66	76	63	57	75	78	91	77	0	0	50	74	78
B. halodurans	57	64	58	58	66	80	89	75	0	0	56	77	76
B. pseudofirmus	57	64	58	60	67	79	87	71	0	0	27	75	78
B. licheniformis	83	29	76	79	81	89	98	89	0	0	29	86	89
B. megaterium	67	81	63	62	74	81	98	79	0	0	26	79	80

Table A.2.2 (Continued)

ORGANISM	SspC	SspD	SspF	YabG	YabP	YabQ	YbaN	YbdM	YcgF	YdfS	YdhD	YerB	YfkQ	YfkR	YfkT
C. sticklandii	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A. metalliredigens	37	46	0	0	38	0	31	0	0	0	0	0	40	22	0
C. difficile	42	43	0	44	32	0	32	0	0	34	0	0	0	0	0
C. phytofermentans	33	40	0	0	38	26	34	0	0	31	37	32	39	0	21
C. saccharolyticum	37	39	28	0	31	28	0	0	0	29	31	0	28	0	0
A. arabaticum	0	0	52	0	0	0	35	0	0	35	0	0	42	27	22
C. botulinum	32	41	0	42	39	25	33	0	0	40	25	0	37	21	21
T. oceani	30	41	50	45	36	0	32	0	25	31	0	37	41	24	0
A. oremlandii	43	51	0	0	0	0	35	0	0	36	31	32	40	21	23
N. thermophilus	34	40	63	0	37	0	0	0	0	34	0	28	38	24	24
S. wolfei	38	40	43	0	30	0	42	0	0	34	0	0	42	24	23
S. lipocalidus	0	0	56	0	36	0	39	0	0	43	0	0	44	25	22
E. eligens	33	43	0	0	32	0	32	0	0	0	0	0	25	31	0
E. rectale	38	37	0	0	34	0	31	0	0	0	0	0	35	0	0
C. bescii	35	41	50	0	34	0	34	0	0	33	0	32	43	23	21
C. kronotskyensis	33	41	50	0	34	0	34	0	0	34	0	32	43	24	21
C. hydrothermalis	33	41	50	0	33	0	36	0	0	35	0	30	43	22	22
C. kristjanssonii	33	41	52	0	33	0	34	0	0	33	0	29	43	23	21
C. saccharolyticus	35	41	52	0	33	0	30	0	25	35	0	34	42	22	21
C. obsidiansis	33	41	52	0	33	0	34	0	0	34	0	29	43	25	20
C. owensensis	33	41	52	0	34	0	34	0	0	33	0	30	43	24	20
R. albus	34	0	0	0	33	0	27	0	0	30	0	0	32	0	0
T. tengcongensis	32	41	61	43	39	28	38	0	0	36	29	0	46	26	23
T. thermosaccharolyticum	31	39	59	40	35	0	37	0	0	33	0	0	43	24	27
Thermoanaerobacter sp. X513	0	35	61	44	39	24	39	0	0	35	30	0	44	27	24
Thermoanaerobacter sp. X514	0	35	61	44	39	24	39	0	0	35	30	0	44	27	24
T. mathranii	31	38	63	42	38	23	40	0	0	35	31	0	44	28	25
T. italicus	0	38	63	42	38	23	40	0	0	35	31	0	44	28	24
T. brockii	0	35	61	44	39	24	39	0	0	35	31	0	44	27	24
T. pseudethanolicus	0	35	61	44	39	24	39	0	0	35	31	0	44	27	24
L. sphaericus	59	74	62	0	42	26	33	36	30	38	33	42	46	25	0
C. novyi	36	40	0	42	29	25	32	0	22	39	35	0	42	20	0
H. orenii	34	34	60	39	31	39	32	0	22	38	30	33	42	24	0
C. thermocellum	27	34	57	44	36	0	34	0	25	38	28	31	37	23	20
C. kluyveri	38	44	40	42	32	0	32	37	26	30	38	0	42	0	0
C. ljungdahlii	40	45	44	43	34	0	33	0	26	40	37	0	42	26	25
B. subtilis	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C. cellulolyticum	35	36	55	36	32	0	33	0	0	39	36	0	43	26	23
E. harbinense	39	35	51	0	38	0	26	0	0	34	33	0	34	22	24
C. beijerinckii	45	43	0	40	35	0	32	0	0	41	24	0	26	0	0
C. perfringens	40	44	0	41	37	23	29	0	25	24	0	31	34	20	0
C. cellulovorans	47	41	0	38	37	0	31	34	0	31	0	0	38	21	0
C. acetobutylicum	43	49	0	39	35	21	29	0	22	41	33	0	40	24	25
C. tetani	0	0	0	41	0	0	36	0	22	39	33	0	43	22	21
P. polymyxa	42	46	74	46	0	28	42	33	54	48	0	31	58	35	26
Paenibacillus sp. JDR-2	43	43	62	43	51	32	44	29	30	38	34	37	51	32	25
Paenibacillus sp. Y412MC10	46	48	72	45	54	0	46	31	52	43	0	37	42	23	26
S. thermophilum	45	42	0	0	38	0	40	0	0	34	34	30	44	23	29
H. modesticaldum	0	0	55	40	35	0	30	0	23	40	0	0	40	23	24
A. acidocaldarius	59	58	72	43	35	0	41	0	0	35	29	0	55	32	28
B. tusciae	34	47	72	46	42	0	39	0	0	37	34	35	43	21	23
Ca. D. audaxviator	39	39	39	40	28	0	40	0	0	42	0	0	36	0	0
A. degensii	35	41	47	43	28	0	45	0	0	43	31	0	43	24	21
M. thermoacetica	0	0	54	41	42	0	32	0	26	43	33	0	47	22	26
T. potens	35	31	63	45	35	0	29	0	28	41	33	37	48	26	27
D. hafniense	47	53	52	43	41	0	41	0	0	38	30	0	42	23	24
D. reducens	42	0	44	39	33	0	39	0	0	53	33	0	43	20	25
P. thermopropionicum	26	30	0	39	34	0	33	0	0	40	31	0	43	26	25
C. hydrogenoformans	35	48	0	46	30	0	43	0	23	39	0	0	45	27	25
D. acetoxidans	0	43	50	41	30	0	31	0	0	37	0	0	42	24	24
O. iheyensis	65	76	76	59	62	33	44	0	51	34	34	46	38	0	0
Geobacillus sp. WCH70	70	76	70	62	71	45	55	0	25	36	34	51	49	30	24
A. flavithermus	66	71	0	66	69	33	55	35	0	35	0	55	42	27	26
B. cellulosilyticus	66	75	71	55	67	41	46	0	51	34	33	45	53	32	26
B. clausii	67	72	70	54	67	0	43	0	55	35	33	0	39	23	22
G. thermodenitrificans	73	0	70	61	79	42	54	34	0	37	68	51	45	26	23
G. kaustophilus	73	72	66	61	71	37	54	35	0	35	68	46	44	26	23
Geobacillus sp. Y4.1MC1	70	75	70	64	71	43	56	34	0	37	34	52	47	30	23
Geobacillus sp. C56-T3	73	74	69	61	71	37	54	35	41	35	68	49	44	26	24
Geobacillus sp. Y412MC52	73	72	67	61	71	37	54	35	22	35	69	50	44	26	25
Geobacillus sp. Y412MC61	73	72	67	61	71	37	54	35	22	35	69	50	44	26	25
B. amyloliquefaciens	75	87	95	92	98	71	80	0	85	55	72	79	45	24	22
B. pumilus	75	81	88	79	73	50	68	0	66	28	32	0	41	0	21
B. brevis	60	65	70	0	59	35	44	31	31	53	34	38	51	35	23
B. weihenstephanensis	70	70	75	59	65	48	53	34	40	34	30	0	38	21	0
B. cereus	65	69	75	59	66	46	53	34	39	35	31	0	38	22	22
B. anthracis	65	70	77	59	70	47	53	34	56	35	31	0	46	22	0
B. thuringiensis	65	70	0	59	72	47	53	34	57	35	31	0	38	22	23
B. halodurans	68	73	70	54	67	42	49	0	57	44	33	42	44	24	0
B. pseudofirmus	62	69	67	56	73	38	50	0	37	56	33	49	40	25	21
B. licheniformis	75	81	83	78	83	57	72	35	68	62	57	64	45	22	24
B. megaterium	66	73	69	64	70	37	55	36	56	41	74	0	46	24	24

Table A.2.2 (Continued)

ORGANISM	YfnG	YfnH	YgaK	YhbH	YhcB	YhcV	YhfW	YisY	YkuD	YkvU	YlaK	YlbB	YlbJ	YmxH	YndD
C. sticklandii	30	0	0	0	0	0	23	0	0	22	0	30	0	0	0
A. metalliredigens	24	0	0	0	0	0	0	23	0	24	0	0	0	0	37
C. difficile	24	25	0	0	25	0	38	43	0	23	0	0	34	51	0
C. phytofermentans	32	0	0	0	0	0	35	46	0	24	0	0	28	32	39
C. saccharolyticum	29	44	37	0	0	0	32	47	0	20	0	0	21	28	25
A. arabaticum	25	0	0	0	30	48	0	0	0	0	51	40	45	35	43
C. botulinum	29	43	48	0	0	45	46	46	60	21	42	38	26	28	41
T. oceani	24	0	0	0	0	0	0	53	0	24	0	0	40	43	47
A. oremlandii	25	0	0	0	0	44	45	49	0	22	0	38	39	46	45
N. thermophilus	25	0	0	0	35	38	44	27	0	22	32	39	45	38	38
S. wolfei	31	0	0	0	0	0	43	27	0	26	29	27	37	35	39
S. lipocalidus	23	0	22	0	0	0	0	0	0	0	0	28	39	36	40
E. eligens	31	0	0	0	0	0	0	0	0	0	0	0	21	0	24
E. rectale	23	0	0	0	0	0	0	25	0	20	0	0	22	31	40
C. bescii	33	0	0	0	0	0	22	0	0	22	0	0	25	32	39
C. kronotskyensis	31	49	0	0	0	0	22	20	0	24	0	0	26	31	39
C. hydrothermalis	32	49	0	0	0	0	21	20	0	22	0	0	28	29	41
C. kristjanssonii	27	0	0	0	0	0	22	20	0	23	0	0	28	29	40
C. saccharolyticus	30	0	0	0	0	0	21	0	0	23	0	0	27	29	40
C. obsidiansis	32	0	0	0	0	0	22	0	0	23	0	0	27	29	40
C. owensensis	27	0	0	0	0	0	22	20	0	24	0	0	28	31	39
R. albus	31	0	0	0	0	0	0	26	0	0	0	0	23	0	30
T. tengcongensis	27	31	0	0	0	0	0	29	0	24	54	0	35	39	48
T. thermosaccharolyticum	35	31	0	0	0	0	0	0	0	23	0	0	38	36	47
Thermoanaerobacter sp. X513	25	31	0	0	0	0	0	0	0	23	0	0	37	36	47
Thermoanaerobacter sp. X514	25	31	0	0	0	0	0	0	0	23	0	0	37	36	47
T. mathranii	27	30	0	0	0	0	0	0	0	23	0	0	37	32	46
T. italicus	35	30	0	0	0	0	0	0	0	23	0	0	37	32	46
T. brockii	29	31	0	0	0	0	0	0	0	23	0	0	37	36	47
T. pseudethanolicus	29	31	0	0	0	0	0	0	0	23	0	0	37	36	47
L. sphaericus	28	29	59	66	0	0	31	26	57	24	0	0	23	31	60
C. novyi	27	28	0	0	0	41	0	0	0	23	39	41	29	32	43
H. orenii	32	0	0	0	0	48	0	27	0	24	49	42	49	34	46
C. thermocellum	34	43	0	35	0	46	44	45	0	23	42	42	41	37	38
C. kluyveri	26	0	0	36	0	45	0	47	61	22	41	36	24	31	44
C. ljungdahlii	29	27	0	35	0	49	44	47	0	22	41	36	25	33	41
B. subtilis	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C. cellulolyticum	29	0	0	37	0	51	44	48	0	0	0	38	35	35	35
E. harbinense	30	0	0	37	0	41	36	26	0	22	0	35	27	36	35
C. beijerinckii	27	0	40	37	28	43	39	48	0	0	0	40	26	31	27
C. perfringens	28	0	33	38	0	0	0	23	0	22	41	0	25	30	33
C. cellulovorans	29	0	0	37	31	0	43	24	0	22	41	0	25	32	40
C. acetobutylicum	27	0	0	36	0	40	40	49	0	22	0	38	25	30	40
C. tetani	26	24	0	37	0	41	43	50	0	22	0	38	25	0	41
P. polymyxa	30	0	49	65	60	42	24	48	58	23	63	46	43	0	56
Paenibacillus sp. JDR-2	34	48	49	65	63	40	51	51	58	31	63	43	43	30	54
Paenibacillus sp. Y412MC10	31	0	42	62	0	0	50	28	60	24	63	0	41	35	61
S. thermophilum	26	24	0	51	0	44	0	25	0	0	0	38	50	43	42
H. modesticaldum	33	31	0	57	0	46	0	0	0	22	55	39	35	40	35
A. acidocaldarius	23	0	0	56	0	47	25	23	57	26	0	37	45	39	43
B. tusciae	30	0	0	63	37	50	20	26	0	23	0	44	44	41	41
Ca. D. audaxviator	28	0	0	50	0	0	0	0	0	24	0	0	44	36	38
A. degensii	32	0	0	43	0	0	0	0	0	24	0	0	39	34	39
M. thermoacetica	29	0	0	56	0	0	0	0	0	28	0	0	45	29	43
T. potens	27	0	0	56	0	0	0	0	0	23	49	30	47	36	41
D. hafniense	30	27	0	54	0	0	40	25	53	23	49	26	40	38	41
D. reducens	52	47	0	56	0	51	0	24	0	26	40	39	44	33	39
P. thermopropionicum	30	28	0	58	0	0	0	0	0	23	32	0	45	36	41
C. hydrogenoformans	27	22	0	51	0	0	0	29	50	26	0	0	33	36	41
D. acetoxidans	31	0	23	55	0	50	0	46	0	25	0	37	45	39	43
O. iheyensis	29	0	0	69	0	0	45	23	0	28	0	0	55	54	39
Geobacillus sp. WCH70	28	27	0	81	38	63	54	24	62	30	0	40	59	78	41
A. flavithermus	27	0	0	78	0	60	50	24	35	29	78	55	62	78	40
B. cellulosilyticus	31	25	0	72	34	48	49	25	60	30	69	49	52	61	41
B. clausii	31	23	0	74	0	0	48	0	66	47	69	0	53	49	40
G. thermodenitrificans	27	0	0	81	0	64	52	0	60	29	77	53	58	78	40
G. kaustophilus	28	30	0	78	0	50	51	0	63	28	77	53	58	79	41
Geobacillus sp. Y4.1MC1	26	0	0	80	38	53	53	26	62	30	77	52	56	79	40
Geobacillus sp. C56-T3	24	0	0	81	0	60	51	0	63	29	77	52	59	79	41
Geobacillus sp. Y412MC52	27	0	0	81	0	50	51	0	63	28	77	52	59	79	40
Geobacillus sp. Y412MC61	27	0	0	81	0	50	51	0	63	28	77	52	59	79	40
B. amyloliquefaciens	91	85	84	94	94	84	68	25	76	92	90	83	79	92	43
B. pumilus	85	74	0	86	84	42	53	27	67	71	80	59	66	88	45
B. brevis	27	43	0	65	0	43	21	44	53	26	68	44	48	42	65
B. weihenstephanensis	30	0	59	80	0	43	55	44	0	31	77	50	53	57	41
B. cereus	33	53	59	80	0	44	55	41	35	32	78	52	54	60	38
B. anthracis	30	0	0	80	0	44	56	42	35	31	78	52	55	59	68
B. thuringiensis	30	0	0	80	0	43	55	43	0	31	77	52	55	59	41
B. halodurans	31	26	0	74	34	50	52	26	48	51	73	48	55	61	45
B. pseudofirmus	28	0	0	73	33	55	51	27	60	54	72	38	58	56	46
B. licheniformis	31	0	0	86	86	75	59	43	75	76	81	61	72	83	63
B. megaterium	30	0	40	79	71	46	49	51	59	51	76	60	55	76	72

Table A.2.2 (Continued)

ORGANISM	YndE	YndF	YngE	YngI	YngJ	YoaR	YobN	YpeB	YqfC	YqfD	YqgT	YqhO	YraD	YrbG	YrkC
C. sticklandii	0	0	39	0	54	0	0	0	0	0	0	0	0	0	0
A. metalliredigens	26	24	0	0	0	0	0	26	0	0	0	0	25	0	0
C. difficile	0	0	0	21	53	0	0	0	33	21	0	0	28	33	0
C. phytofermentans	0	0	28	23	0	0	29	0	0	24	0	0	0	22	0
C. saccharolyticum	0	0	30	37	51	0	29	0	28	29	0	26	0	28	55
A. arabaticum	25	26	0	60	0	0	0	27	34	23	30	26	0	40	66
C. botulinum	28	25	0	39	48	0	25	28	33	0	0	36	0	35	60
T. oceani	29	25	41	0	0	0	27	25	35	27	24	27	0	30	0
A. oremlandii	26	21	38	0	56	0	0	22	42	24	0	0	31	33	0
N. thermophilus	23	21	39	0	52	0	0	0	36	0	30	23	0	36	0
S. wolfei	24	23	0	34	48	0	0	0	36	21	0	25	27	35	0
S. lipocalidus	24	26	0	37	48	36	0	0	38	23	0	26	26	35	0
E. eligens	0	0	30	29	0	0	0	0	37	0	0	0	0	0	0
E. rectale	0	0	27	24	48	0	0	0	32	0	0	0	0	0	0
C. bescii	27	23	40	58	41	0	0	29	35	21	0	27	0	31	0
C. kronotskyensis	27	23	40	58	42	0	0	29	33	21	0	28	0	31	0
C. hydrothermalis	27	21	40	57	42	0	0	28	32	0	0	28	0	31	0
C. kristjanssonii	27	20	40	58	42	0	0	28	0	20	0	26	0	31	0
C. saccharolyticus	28	24	39	58	43	0	0	27	27	0	0	27	0	29	0
C. obsidiansis	29	20	39	59	42	0	0	28	0	0	0	28	0	30	0
C. owensensis	27	22	39	58	42	0	0	28	32	22	0	28	0	31	0
R. albus	0	0	29	30	0	0	0	24	0	26	0	0	0	31	0
T. tengcongensis	30	24	42	33	53	0	0	30	35	25	0	0	0	33	0
T. thermosaccharolyticum	30	26	40	51	55	0	0	32	31	26	0	0	0	35	0
Thermoanaerobacter sp. X513	29	24	41	32	0	0	0	30	34	24	0	0	0	32	0
Thermoanaerobacter sp. X514	29	24	41	32	0	0	0	30	34	24	0	0	0	32	0
T. mathranii	28	23	41	32	0	0	0	31	30	24	0	0	0	33	0
T. italicus	28	23	41	32	0	0	0	31	30	24	0	0	0	33	0
T. brockii	29	24	41	32	0	0	0	30	34	24	0	0	0	32	0
T. pseudethanolicus	29	24	41	32	0	0	0	30	34	24	0	0	0	32	0
L. sphaericus	61	42	74	61	76	0	0	23	0	23	42	55	0	46	0
C. novyi	25	26	0	0	51	0	0	25	34	0	0	41	0	31	61
H. orenii	29	25	0	0	0	0	0	28	34	26	0	25	32	34	0
C. thermocellum	22	26	40	0	0	0	0	26	29	26	0	29	29	34	60
C. kluyveri	0	0	21	29	0	0	0	23	30	28	33	38	26	32	58
C. ljungdahlii	0	27	0	30	42	0	0	25	28	0	33	34	51	28	62
B. subtilis	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C. cellulolyticum	21	21	39	57	38	0	30	28	28	26	0	30	0	33	69
E. harbinense	24	25	0	26	36	0	0	24	23	25	30	0	0	34	0
C. beijerinckii	0	0	0	29	49	0	27	0	35	23	0	0	30	29	65
C. perfringens	0	0	0	0	52	0	0	0	31	24	0	0	0	29	0
C. cellulovorans	20	23	0	35	50	0	29	0	34	25	0	37	29	37	0
C. acetobutylicum	22	27	22	0	50	0	0	23	34	27	0	0	47	35	0
C. tetani	24	22	0	0	52	0	0	26	0	0	0	0	0	36	58
P. polymyxa	50	39	0	28	0	0	0	36	41	31	36	40	0	36	0
Paenibacillus sp. JDR-2	45	34	0	34	41	54	28	41	42	32	44	40	0	42	67
Paenibacillus sp. Y412MC10	53	43	0	35	0	57	28	43	38	30	40	42	0	40	53
S. thermophilum	22	21	69	34	63	0	0	26	32	0	41	33	33	25	63
H. modesticaldum	23	21	39	61	51	0	0	27	39	26	0	0	0	41	0
A. acidocaldarius	21	23	71	35	65	0	0	0	32	26	0	25	0	28	0
B. tusciae	25	20	73	34	67	0	0	41	25	23	0	25	0	38	0
Ca. D. audaxviator	0	0	26	25	0	0	0	25	39	23	0	0	0	32	0
A. degensii	22	25	0	23	0	0	0	0	36	26	0	0	0	37	0
M. thermoacetica	25	28	40	30	0	0	0	26	37	21	0	0	0	31	0
T. potens	26	28	0	63	0	0	0	32	38	23	0	40	0	36	0
D. hafniense	24	25	0	41	53	0	0	23	35	23	0	28	28	34	0
D. reducens	24	23	0	33	52	0	0	30	37	0	0	27	0	38	0
P. thermopropionicum	24	23	38	60	53	0	0	26	35	0	0	33	0	39	0
C. hydrogenoformans	23	23	41	34	54	0	0	26	27	22	0	24	29	31	0
D. acetoxidans	24	26	36	39	53	0	0	27	36	24	0	0	0	38	0
O. iheyensis	23	23	74	62	76	50	41	46	42	34	48	46	31	28	0
Geobacillus sp. WCH70	28	24	79	36	79	0	0	56	64	46	40	56	0	31	73
A. flavithermus	26	25	79	66	79	0	0	58	61	42	50	59	0	60	75
B. cellulosilyticus	26	21	40	33	50	57	0	50	55	39	40	47	0	52	56
B. clausii	26	26	74	65	76	0	0	41	58	40	0	53	0	35	53
G. thermodenitrificans	25	23	77	67	78	0	54	53	63	43	48	54	45	34	64
G. kaustophilus	25	25	77	66	79	0	0	54	62	45	47	53	53	0	65
Geobacillus sp. Y4.1MC1	26	25	78	67	77	0	56	56	63	45	45	56	0	31	0
Geobacillus sp. C56-T3	25	25	76	67	78	0	56	54	62	45	47	54	51	0	0
Geobacillus sp. Y412MC52	26	23	77	67	78	0	56	54	62	45	47	54	53	0	0
Geobacillus sp. Y412MC61	26	23	77	67	78	0	56	54	62	45	47	54	53	0	0
B. amyloliquefaciens	36	32	84	73	87	0	0	81	94	77	62	84	81	70	82
B. pumilus	34	28	79	61	80	0	0	66	87	65	47	65	0	60	0
B. brevis	53	50	70	59	61	52	50	42	51	32	0	43	63	40	60
B. weihenstephanensis	26	27	76	33	75	31	68	58	61	46	0	40	0	50	71
B. cereus	58	50	76	34	51	30	68	57	61	45	0	40	0	51	0
B. anthracis	49	51	76	34	75	31	69	57	62	46	0	40	0	51	0
B. thuringiensis	27	27	76	34	75	31	69	57	0	46	0	40	0	51	0
B. halodurans	25	22	78	66	82	0	0	42	60	41	38	51	43	49	0
B. pseudofirmus	33	27	79	68	80	51	41	44	62	40	42	54	35	49	0
B. licheniformis	58	55	82	71	82	75	0	73	86	69	59	72	50	62	72
B. megaterium	58	49	72	56	63	56	0	58	60	48	50	61	73	59	62

Table A.2.2 (Continued)

ORGANISM	YtcA	YtcC	YtfJ	YtlA	YtlC	YtlD	YtvI	YtxC	YunB	YutH	ywjD	YyaC	YyaE	YyaO	YybI
C. sticklandii	26	0	0	23	0	26	26	0	0	0	0	0	33	0	0
A. metalliredigens	27	25	0	48	48	45	24	0	29	0	0	43	26	0	0
C. difficile	39	0	41	0	38	44	28	0	25	0	0	41	31	0	0
C. phytofermentans	35	0	27	46	47	42	28	0	0	0	0	45	0	0	0
C. saccharolyticum	0	0	21	46	44	43	27	0	0	0	0	44	0	0	0
A. arabaticum	39	23	48	47	46	44	27	0	25	0	0	0	27	0	41
C. botulinum	26	28	45	45	46	50	25	0	30	22	0	52	0	0	50
T. oceani	38	25	53	21	35	27	29	31	31	0	0	60	35	0	0
A. oremlandii	0	0	49	46	46	45	26	24	28	0	0	0	38	0	0
N. thermophilus	42	0	47	0	0	0	21	0	0	0	0	55	37	0	0
S. wolfei	28	0	48	0	0	0	29	0	24	0	0	50	22	0	0
S. lipocalidus	0	0	49	0	36	23	0	26	26	0	0	0	23	0	0
E. eligens	0	0	25	46	43	39	0	0	0	0	0	0	0	0	0
E. rectale	0	20	26	0	0	0	0	0	0	0	0	0	0	0	0
C. bescii	26	0	48	0	39	26	0	0	0	0	0	0	35	0	0
C. kronotskyensis	26	0	49	22	38	26	0	0	0	0	0	0	35	0	0
C. hydrothermalis	26	0	49	21	41	26	0	0	0	0	0	0	35	0	0
C. kristjanssonii	39	0	49	20	39	25	22	0	0	0	0	0	35	0	0
C. saccharolyticus	26	0	47	22	38	28	26	25	0	0	0	0	35	0	0
C. obsidiansis	26	0	49	0	0	0	21	0	0	0	0	0	36	0	0
C. owensensis	26	0	48	0	0	0	22	0	0	0	0	0	35	0	0
R. albus	28	0	39	0	39	24	21	0	0	0	0	0	0	0	0
T. tengcongensis	25	0	52	47	48	49	28	0	24	0	0	54	0	0	0
T. thermosaccharolyticum	0	0	54	42	45	48	31	0	25	0	0	59	0	0	0
Thermoanaerobacter sp. X513	38	0	48	44	50	49	29	0	26	0	0	57	0	0	0
Thermoanaerobacter sp. X514	38	0	48	44	50	49	29	0	26	0	27	57	0	0	0
T. mathranii	0	0	49	44	50	48	29	0	26	0	0	56	0	0	0
T. italicus	0	0	48	44	51	48	29	0	26	0	0	56	0	0	0
T. brockii	0	0	49	44	50	49	29	0	26	0	0	57	0	0	0
T. pseudethanolicus	0	0	49	44	50	49	29	0	26	0	0	57	0	0	0
L. sphaericus	26	0	0	63	60	69	28	0	29	0	0	0	0	0	55
C. novyi	26	0	45	45	48	47	21	22	0	21	0	0	0	0	0
H. orenii	41	0	46	46	49	45	31	27	27	0	36	49	0	0	0
C. thermocellum	40	42	52	48	47	47	29	0	28	22	40	50	0	0	36
C. kluyveri	43	0	47	40	47	47	27	23	0	22	0	54	0	77	0
C. ljungdahlii	28	25	46	42	48	49	27	24	29	21	0	53	34	0	37
B. subtilis	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C. cellulolyticum	27	23	45	48	45	48	28	23	0	0	40	0	0	0	56
E. harbinense	29	22	36	29	46	34	24	0	26	0	0	42	0	0	0
C. beijerinckii	38	27	43	43	47	46	0	21	25	0	34	46	23	0	0
C. perfringens	38	24	42	0	0	0	29	23	23	0	0	50	32	0	26
C. cellulovorans	26	0	43	24	38	0	29	22	25	0	30	51	0	0	0
C. acetobutylicum	28	0	44	43	43	46	0	0	23	0	0	49	0	0	0
C. tetani	27	26	42	44	40	44	25	0	26	0	0	53	0	0	51
P. polymyxa	41	27	56	58	50	57	34	24	25	0	39	51	33	0	0
Paenibacillus sp. JDR-2	41	27	55	59	49	56	27	0	25	0	39	57	26	0	0
Paenibacillus sp. Y412MC10	26	30	61	61	51	53	37	28	23	0	43	52	33	0	0
S. thermophilum	38	40	45	37	46	45	27	0	23	0	0	56	32	0	0
H. modesticaldum	23	0	50	0	0	0	34	27	27	0	0	49	23	0	0
A. acidocaldarius	28	0	64	0	0	25	23	0	0	0	0	54	38	0	0
B. tusciae	24	24	53	38	46	43	29	0	25	0	0	52	24	0	0
Ca. D. audaxviator	0	24	44	28	34	0	33	0	25	0	0	0	26	0	0
A. degensii	0	0	54	22	0	20	31	0	21	0	0	52	23	0	0
M. thermoacetica	0	0	48	25	37	30	30	30	23	0	0	51	27	0	0
T. potens	25	0	54	24	38	28	28	0	29	0	0	52	28	0	0
D. hafniense	37	24	49	46	50	49	30	24	0	0	0	51	35	0	0
D. reducens	36	0	53	51	45	49	33	25	26	0	0	52	37	0	0
P. thermopropionicum	29	0	52	32	40	31	33	27	26	0	0	54	29	0	0
C. hydrogenoformans	0	30	53	0	35	0	29	21	22	0	0	56	38	0	0
D. acetoxidans	38	42	43	29	38	0	32	0	22	0	0	51	23	0	53
O. iheyensis	39	0	57	61	59	63	35	25	29	30	0	48	0	0	29
Geobacillus sp. WCH70	37	0	71	68	65	66	48	40	46	43	0	56	0	0	0
A. flavithermus	48	54	71	67	60	67	51	36	47	39	0	0	0	0	0
B. cellulosilyticus	27	0	72	60	57	69	41	0	34	0	37	49	26	0	0
B. clausii	41	0	66	59	55	66	30	34	33	27	0	42	27	0	0
G. thermodenitrificans	54	54	61	67	61	64	50	42	44	41	33	56	33	0	0
G. kaustophilus	54	55	65	68	63	62	48	36	44	40	0	55	60	0	0
Geobacillus sp. Y4.1MC1	38	44	68	67	66	66	47	45	46	41	0	56	62	0	0
Geobacillus sp. C56-T3	54	55	65	67	63	63	48	36	44	40	0	57	61	0	0
Geobacillus sp. Y412MC52	54	55	67	0	63	62	48	36	44	40	0	57	61	0	0
Geobacillus sp. Y412MC61	54	55	67	0	63	62	48	36	44	40	0	57	61	0	0
B. amyloliquefaciens	40	0	87	76	85	81	77	75	79	77	77	83	83	0	0
B. pumilus	40	0	76	66	73	73	54	51	56	60	68	67	33	0	0
B. brevis	41	37	64	60	55	57	36	0	29	22	38	0	31	0	30
B. weihenstephanensis	54	56	57	66	56	64	43	32	39	43	58	53	0	0	0
B. cereus	39	26	55	66	56	64	43	34	38	41	57	53	0	0	0
B. anthracis	38	27	57	66	56	64	44	31	38	41	57	55	0	0	0
B. thuringiensis	39	0	57	65	56	64	44	31	38	41	57	55	0	0	0
B. halodurans	29	39	63	0	38	27	38	33	37	33	32	47	31	0	0
B. pseudofirmus	36	0	68	65	61	67	32	32	38	32	35	0	33	0	51
B. licheniformis	39	25	80	69	72	77	61	63	58	65	0	71	31	96	0
B. megaterium	42	0	75	65	63	76	45	35	41	43	61	57	25	0	0